

Eigentijds toetsen en beoordelen

De opbrengst van vier jaar praktijkonderzoek

**Desirée Joosten-ten Brinke, Jorik Arts, Carel Burghout, Mieke Jaspers,
Wim Lansu, Monique Maas, Jos Verhoosel en Suzanne Kratsborn**

Tilburg, Fontys Lerarenopleiding Tilburg

Juni 2014

Eigentijds toetsen en beoordelen

De opbrengst van vier jaar praktijkonderzoek

Desirée Joosten-ten Brinke, Jorik Arts, Carel Burghout, Mieke Jaspers,
Wim Lansu, Monique Maas, Jos Verhoosel en Suzanne Kratsborn

Tilburg, Fontys Lerarenopleiding Tilburg
Juni 2014

Colofon

Deze praktijkpublicatie is een product van het lectoraat 'Eigentijds toetsen en beoordelen' van de Fontys Lerarenopleiding Tilburg

Auteurs

Desirée Joosten-ten Brinke, Jorik Arts, Carel Burghout, Mieke Jaspers, Wim Lansu, Monique Maas, Jos Verhoosel en Suzanne Kratsborn

Opmaak: Grafische Producties, Fontys Hogescholen

2014 © Fontys Lerarenopleiding Tilburg

Copyright

Alles uit deze uitgave mag gekopieerd, gedistribueerd, vertoond en opgevoerd worden, mits het werk in de originele staat blijft, de auteursnaam vermeld wordt en niet voor commerciële doeleinden gebruikt wordt.

Lees voor gebruik de volledige licentie op de [website](#).



Inhoudsopgave

Voorwoord		5	4.4	Onderzoeksvraag	41
			4.5	Beoogde resultaten	41
Hoofdstuk 1: Eigentijds toetsen en beoordelen		7	4.6	Methode: twee pilot cases	41
1.1	Inleiding	7	4.7	Resultaten	43
1.2	De kwaliteitspiramide voor eigentijds toetsen en beoordelen	7	4.8	Conclusies en discussie	44
			Hoofdstuk 5: Praktijkonderzoek naar inzet van rubrics, checklists en scripts bij vakdidactiek		47
Hoofdstuk 2: Kwaliteit van schriftelijke feedback		11	5.1	Samenvatting	47
2.1	Samenvatting	11	5.2	Inleiding	47
2.2	Inleiding	11	5.3	Onderzoeksvraag	48
2.3	Onderzoeksvragen	13	5.4	Methode	49
2.4	Methode	13	5.5	Resultaten	51
2.5	Resultaten	15	5.6	Conclusies en discussie	52
2.6	Discussie & Conclusies	17			
Hoofdstuk 3: De meerwaarde van een vormgever		23	Hoofdstuk 6: De invloed van Random Assessment met Peer-Assessment (RAPA) op studiegedrag		55
3.1	Samenvatting	23	6.1	Samenvatting	55
3.2	Inleiding	23	6.2	Inleiding	55
3.3	Onderzoeksvragen	27	6.3	Onderzoeksvragen	59
3.4	Methode	27	6.4	Methode	59
3.5	Resultaten	28	6.5	Resultaten	61
3.6	Conclusie en discussie	32	6.6	Conclusie en discussie	64
Hoofdstuk 4: Effecten en resultaten van het inzetten van rubrics		37	6.7	Aanbevelingen	65
4.1	Samenvatting	37	Bijlage 6.1 Vragenlijst LOB-studenten deeltijd		66
4.2	Inleiding	37			
4.3	De verlegenheidssituaties	40	Bijlage 6.2 De rubric peer-assessment LOB*		70

Hoofdstuk 7: De landelijke kennistoets Geschiedenis	73	Hoofdstuk 10: Tot slot	99
7.1 Vooraf	73	10.1 Bijdrage aan de professionalisering van leraren en lerarenopleiders	101
7.2 Samenvatting	73	10.2 De professionele praktijk	102
7.3 Wie is betrokken bij een landelijke toets?	74	10.3 Het vervolg	102
7.4 Van kennisbasis naar kennistoets	74		
7.5 De pretest en domeintoets: zicht op de kwaliteit van items	75	Literatuur	103
7.6 De ervaringen van studenten	77		
7.7 Conclusie	77	CV van de lectoraatsleden	109
Hoofdstuk 8: Trainen van beoordelingsvaardigheden	79	Bijlage 1: Overzicht van workshops, presentaties en publicaties in 2010-2014	113
8.1 Samenvatting	79		
8.2 Inleiding	79		
8.3 Onderzoeksvraag	80		
8.4 Methoden	81		
8.5 Resultaten	82		
8.6 Conclusie	85		
8.7 Discussie	88		
Hoofdstuk 9: Bewust en bekwaam toetsen en beoordelen	89		
9.1 Samenvatting	89		
9.2 Aanleiding en probleemstelling	90		
9.3 Het ontwerp en de ontwikkeling van het zelfbeoordelingsinstrument	92		
9.4 Resultaat en conclusie van het project	97		

Voorwoord

Voor u ligt een praktijkpublicatie van het lectoraat 'Eigentijds toetsen en beoordelen'. Dit lectoraat is in 2010 gestart. In deze publicatie wordt de opbrengst van het lectoraat gepresenteerd aan de hand van inhoudelijke bijdragen.

Het thema van het lectoraat is *Toetsen en beoordelen*. Voor lerarenopleiders en docenten cruciale onderdelen van hun beroep. Het realiseren van kwaliteit van toetsen en beoordelen is ingewikkeld. De kwaliteit van een losse toets kan op orde zijn, maar als de toets niet aansluit bij de doelen van een vak, als deze doelen niet goed geformuleerd zijn of als er geen afstemming tussen toetsen in een toetsprogramma is, is er uiteindelijk toch sprake van een zwak systeem van toetsen. Het belang van goed toetsen is de afgelopen jaren steeds duidelijker geworden.

In het lectoraat hebben kenniskringleden onderzoek gedaan naar hun eigen praktijk om daarmee zichzelf te professionaliseren in praktijkonderzoek én in het thema van het lectoraat. Een belangrijke focus was het formatief toetsen, waarbij onderzocht is in hoeverre

formatieve methoden effectief in het onderwijs kunnen worden ingezet. Achtereenvolgens worden in deze publicatie onderzoeken gepresenteerd over het geven van schriftelijke feedback, het werken met rubrics, scripts en checklists, het beoordelen van groepswerk, peer assessment, de kwaliteit van kennisvragen bij *10voordeleraar* en toetsbekwaamheid van lerarenopleiders.

Na elk hoofdstuk geven de kenniskringleden aan wat het doen van praktijkonderzoek voor hen als persoon en als lerarenopleider betekent.

We hopen dat deze praktijkpublicatie u ideeën en tips geeft om verder te investeren in kwaliteit van toetsing.

Desirée Joosten-ten Brinke

Samenstelling lectoraat 2010-2014

De samenstelling van het lectoraat is met een vaste kerngroep redelijk constant geweest. Hieronder een overzicht van de personen die in de eerste periode lid zijn (geweest).

Naam	Periode	Opleiding/afdeling
Desirée Joosten-ten Brinke	2010-2014	
Jorik Arts	2010-2014	Biologie
Carel Burghout	2010-2014	Engels
Mieke Jaspers	2010-2014	Onderwijs en Onderzoek
Jos Verhoosel	2010-2014	Wiskunde
Monique Maas	2012-2014	Engels
Wim Lansu	2010-2013	Geschiedenis
Kathi Barth	2011-2012	Engels
Hjalmar Janson	2010-2011	Economie
Suzanne Kratsborn	2011-2012	Pabo Tilburg
Mara Rooijmans	2011	Stedelijk College Eindhoven

Hoofdstuk 1

Eigentijds toetsen en beoordelen

1.1 Inleiding

Er wordt in het onderwijs geïnvesteerd in de kwaliteit van toetsing. Het rapport van Commissie Bruijn (2012) geeft concrete aanbevelingen om de kwaliteit van het beoordelen in het hoger onderwijs te verbeteren. Examencommissies hebben een proactieve rol gekregen als gevolg van de Wet Versterking Besturing (2010). En de standaard in de beperkte opleidingsaccreditatie die direct gerelateerd is aan toetsing zal zelfs in de komende periode uitgesplitst worden in twee afzonderlijke standaarden (NVAO, 2014). Naast het realiseren van toetskwaliteit hebben onderwijsinstellingen ook als doelstelling het behalen van een zo hoog mogelijk rendement. Kwaliteit van beoordelen is ook hierbij essentieel (zie onder andere Cohen-Schotanus, 2010; Schmidt, 2012). Opleidingen met een goede combinatie van toetsing, waarbij het geven van feedback essentieel is, halen relatief hoge rendementen (Gibbs, 2010). Verkeerde beoordelingen kunnen leiden tot een lager rendement, maar erger nog, tot het onterecht slagen of zakken van studenten. Kwaliteit van toetsing vraagt om deskundige docenten.

Investeren in toetskwaliteit is belangrijk, aangezien op basis van toetsing kleine en grote beslissingen over de voortgang van studenten worden genomen. Toetsing is een sterk sturende factor in het bepalen van kwaliteit van het onderwijs. Omdat toetsing niet los staat van het onderwijs en de kwaliteit van de toetsing steeds meer moet worden verantwoord, zal de opleider zich moeten professionaliseren in wetenschappelijke, maatschappelijke en praktijkgerichte ontwikkeling van toetsing. Om kwaliteit te realiseren en te kunnen waarborgen moet het voor opleidingen duidelijk zijn hoe de gewenste kwaliteit kan worden bereikt (Van Berkel, Sluijsmans, Joosten-ten Brinke, in voorbereiding).

1.2 De kwaliteitspiramide voor eigentijds toetsen en beoordelen

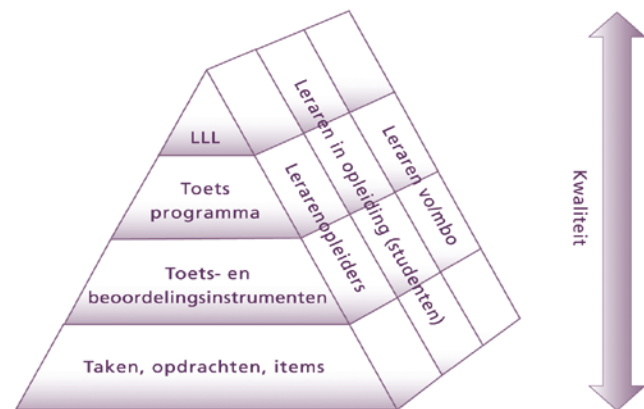
Wat houdt toetskwaliteit in en wanneer is het goed genoeg? Om hier uitspraken over te kunnen doen, is in de eerste lectoraatsperiode de kwaliteitspiramide voor eigentijds toetsen en beoordelen ontworpen (zie figuur 1.1; Joosten-ten Brinke, 2010).

In de kwaliteitspiramide worden vier lagen onderscheiden:

- Kwaliteit van het toetsbeleid in de context van leven lang leren (LLL): hiermee wordt de kwaliteit van de toetsing vanuit de visie van de opleiding bedoeld. In de piramide wordt dit aangeduid met LLL, om aan te geven dat toetsing zinvol moet zijn voor de lange termijn. Kwaliteit van toetsing begint bij de visie die opleidingen hebben ten aanzien van toetsen en beoordelen. Het inrichten van de toetsing is een ontwerpvragestuk, waarbij de aansluiting met het leren essentieel is. In hoeverre draagt het toetsprogramma bij aan een duurzaam leerresultaat? Hoe is de toetsing verbonden aan kennis en vaardigheden voor de lange termijn?
- Kwaliteit van toetsprogramma's: hiermee wordt de kwaliteit van het toetsprogramma als geheel bedoeld. Het gaat om de heldere samenhang tussen toetsen en de borging van de verschillende functies van toetsing.
- Kwaliteit van toetsen: hiermee wordt de kwaliteit van de hele toets met bijbehorende beoordelingsinstrumenten bedoeld, inclusief de voorbereiding, uitvoering en terugkoppeling. Ook gaat het hier om de kwaliteit van de samenhang tussen de items. Voorbeelden van kwaliteitscriteria zijn validiteit, betrouwbaarheid en bruikbaarheid;
- Kwaliteit van toetstaken en toetsitems: elke taak, opdracht of vraag in een toets moet voldoen aan de criteria relevantie, objectiviteit, efficiency en moeilijkheid.

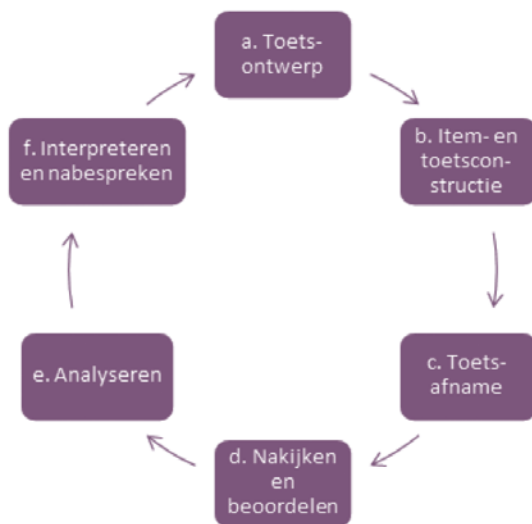
De niveaus staan niet op zichzelf. Kwaliteit van toetsing wordt geborgd door een sterke samenhang van de vier lagen. Is de kwaliteit op de onderste laag niet goed, dan heeft dit negatieve invloed op de kwaliteit van de bovenliggende lagen. Andersom, idem dito. Als onduidelijk is wat de visie op toetsing is, zal dit zijn uitwerking naar beneden hebben op de kwaliteit van toetsprogramma's, toetsen en de onderliggende items.

Voor lerarenopleidingen is de kwaliteit van toetsen en beoordelen voor verschillende doelgroepen van belang. De studenten worden opgeleid tot docenten en de lerarenopleider staat model voor goed onderwijs voor de studenten. Dit betekent dat het lectoraat zich richt op de groepen lerarenopleiders, docenten en studenten (zie figuur 1.1).



Figuur 1.1 De kwaliteitspiramide voor eigentijds toetsen en beoordelen rekening houdend met de doelgroepen lerarenopleiders, docenten in opleiding en docenten in het werkveld.

Om kwaliteit te realiseren is het belangrijk dat het toetsproces volledig doorlopen wordt (zie figuur 1.2). In de ontwerpfase zullen vragen over de functie van toetsen, de keuze voor toetsvormen, de beoordelaars, de uitleervorm belangrijk zijn, terwijl bij de fase van item- en toetsconstructie deskundigheid nodig is in het beoordelen van de kwaliteit van de geconstrueerde vragen en samengestelde toetsen. Voor het borgen van de kwaliteit is het belangrijk dat het hele proces doorlopen wordt. Na een toetsafname is analyse nodig van toetsen en de toetsvragen en levert interpretatie van de resultaten informatie op voor studenten en docenten. Deze informatie kan met feedback studenten inzicht geven in hun toetsresultaten zodat zij daarvan kunnen leren voor een toekomstige situatie. Docenten nemen de informatie mee naar een volgend toetsontwerp.



Figuur 1.2 Het toetsproces

Onderzoek laat zien dat onderwijs waarbij de nadruk ligt op het summatieve toetsen - het toetsen om beslissingen te nemen - het leren van leerlingen kan belemmeren (Black & Wiliam, 1998). De afgelopen decennia is er daardoor steeds meer aandacht voor toetsen die het leren juist stimuleren. Deze toetsing wordt het formatief toetsen genoemd. Formatief toetsen is echter pas effectief voor leren als er een goede afstemming is tussen condities waaronder formatief toetsen wordt vormgegeven, de doelen van formatief toetsen, keuze van methoden en de te verwachten uitkomsten op leren van leerlingen (Sluijsmans, Joosten-ten Brinke, & Van der Vleuten, 2012). Een van de condities daarbij is dat docenten en lerarenopleiders in staat zijn om kennis over formatief toetsen op te bouwen en te vertalen naar een visie en implementatie naar de eigen onderwijspraktijk.

In het lectoraat Eigentijds toetsen en beoordelen zijn verschillende onderzoeken uitgevoerd om een bijdrage te leveren aan het creëren van een gezamenlijk bewustzijn van het belang van goede toetsing en beoordeling en een verdere professionalisering van lerarenopleiders en docenten op dit domein.

De volgende hoofdstukken beschrijven verschillende onderzoeken die uitgevoerd zijn in de eerste lectoraatsperiode. Deze onderzoeken dragen bij aan het streven van het lectoraat naar een kwaliteitsslag in toetsbewustzijn en toetsbekwaamheid bij docenten en studenten.

Achtereenvolgens worden de volgende onderzoeken beschreven:

1. Kwaliteit van schriftelijke feedback
2. De meerwaarde van een vormgever voor het geven van feedback
3. Effecten en resultaten van het inzetten van rubrics
4. Praktijkonderzoek naar inzet van rubrics, checklists en scripts bij vakdidactiek
5. De invloed van Random Assessment met Peer-Assessment (RAPA) op het studiegedrag van deeltijdstudenten wiskunde
6. De landelijke kennistoets Geschiedenis
7. Trainen van beoordelingsvaardigheden
8. Kennis en kunde van lerarenopleiders



Hoofdstuk 2

Kwaliteit van schriftelijke feedback

Jorik Arts en Mieke Jaspers

2.1 Samenvatting

Feedback vormt een belangrijk onderdeel van een effectieve formatieve toetspraktijk. Diverse publicaties beschrijven het sterke effect op leren dat door middel van feedback gerealiseerd kan worden. Feedback leidt echter niet zonder meer tot leren, het kan zelfs een negatief effect hebben op leren. In de literatuur worden diverse criteria beschreven die bepalend zijn voor de effectiviteit van feedback. In deze studie wordt onderzocht in hoeverre de feedbackpraktijk in de onderzochte context voldoet aan de succescriteria uit de literatuur. Hierbij is er specifiek gekeken naar de schriftelijke feedback. De schriftelijke feedback is geanalyseerd en de perceptie van de studenten is onderzocht aan de hand van een vragenlijst. De feedbackpraktijk blijkt te voldoen aan meerdere succescriteria, er worden echter ook verbeterpunten gevonden. Deze hebben betrekking op het gebruik van feed up en feed forward en op het gebruik van feedback op proces- en zelfregulatie-niveau. Uitgaande van de inzichten uit de literatuur zou de kwaliteit van leren toenemen bij meer aandacht voor voorgenoemde vormen van feedback.

2.2 Inleiding

Feedback vormt een essentieel bestanddeel van een effectieve (formatieve) toetspraktijk. Als omschrijving van feedback wordt, in navolging van de meeste literatuur op dit gebied, uitgegaan van de definitie van Ramaprasad (1983). Deze definitie luidt: *"Feedback is information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way"*. Om feedback tot een leereffect te laten leiden is het echter noodzakelijk dat er enkele richtlijnen nageleefd worden (zie onder andere Gibbs & Simpson, 2004; Nicol & Macfarlane-Dick, 2006). Zo is er binnen een goede feedbackpraktijk in elk geval aandacht voor de volgende aspecten:

- Feedback schept helderheid over doelen, criteria en verwacht eindniveau;
- Feedback maakt duidelijk *hoe* het verwacht niveau bereikt kan worden;
- Feedback is helder geformuleerd en volledig;
- Feedback ondersteunt motivatie en zelfvertrouwen.

Deze aanbevelingen komen ook terug in het model van Hattie & Timperley (2007). Zij geven aan dat effectieve

feedback ingedeeld kan worden in drie categorieën: *feed up*, *feedback* en *feed forward*. Feed up vormt het antwoord op de vraag: “waar werk ik naar toe?” en refereert daarmee naar de leerdoelen. Feedback heeft betrekking op alle opmerkingen die te maken hebben met de tot dan toe geleverde prestatie. Het geeft daarmee antwoord op de vraag: “hoe doe ik het tot nu toe?”. Ten slotte moet er aandacht zijn voor ontwikkelpunten voor de student voor toekomstig werk. Dit wordt in het model van Hattie en Timperley (2007) aangeduid als *feed forward* en geeft antwoord op de vraag “hoe nu verder?”.

Bovengenoemde categorieën kunnen volgens Hattie en Timperley (2007) ieder nog opgedeeld worden in verschillende niveaus: taak, proces, zelfregulatie en persoon. Feedback op taakniveau geeft aan of een taak juist is uitgevoerd of niet. Daarbij kan gedacht worden aan een opmerking als “literatuurverwijzing is niet volgens APA richtlijnen”. Feedback op procesniveau geeft aan welke aandachtspunten er zijn met betrekking tot de gekozen aanpak of strategie, bijvoorbeeld: “maak ook eens gebruik van Engelstalige zoektermen”. Feedback op het niveau van zelfregulatie heeft (onder meer) betrekking op zelf-evaluatie. Denk daarbij aan opmerkingen als: “wat zijn nu de sterke en minder sterke punten van dit werkstuk?”. Feedback kan tot slot ook nog gericht zijn op de persoon (“wat ben je toch een creatieve student”). Uit de meta-analyse van Hattie en Timperley (2007) komt naar voren dat met name feedback op het niveau van proces en zelfregulatie tot leeropbrengst leidt. Feedback op het niveau van taak kan tot een leereffect leiden wanneer de student ook begrijpt *waarom* iets juist is

uitgevoerd of niet. Een opmerking als “goed”, “prima” of “literatuurverwijzing!” zal niet automatisch tot een leereffect leiden wanneer het onduidelijk is wat er nu eigenlijk goed is, of wat er nu eigenlijk mis is met de literatuurverwijzingen. Dit soort opmerkingen moeten gepaard gaan met een toelichting. Hetzelfde geldt eigenlijk voor feedback op persoon. Als er een uitspraak gedaan wordt over de creativiteit van de student, waar blijkt die creativiteit dan uit?

De uitleg bij feedback is een aspect dat meegenomen is in een analyse van Glover en Brown (2006). Zij duiden dit aan met diepte 1, 2 en 3. Waarbij diepte 1 bestaat uit een indicatie (“uitstekend”), diepte 2 uit een correctie (“deze alinea verplaatsen naar de inleiding”) en diepte 3 uit een indicatie of correctie gepaard met een uitleg. In het onderzoek van Glover en Brown komt naar voren dat er weinig aandacht lijkt voor uitleg (diepte 3) bij schriftelijke feedback op werkstukken van studenten. Iets wat vervolgens ook naar voren komt bij een studie door Walker (2009). Wanneer een student de feedback niet begrijpt dan zal er ook niets zinnigs mee gedaan kunnen worden. Terugkijkend naar de definitie van feedback blijft de ‘gap’ bestaan. Een indicatie geeft aan dat er blijkbaar aan een bepaald doel is voldaan (of juist niet), maar laat in het midden hoe de student zich verder kan (of moet) ontwikkelen. Een verbetering geeft indirect informatie over de gewenste prestatie, maar het denkwerk is al door de docent ingevuld. De student kan dan iets aanpassen zonder dat er iets geleerd is.

Vanuit het voorgaande kan voorzichtig geconcludeerd worden dat feedback niet zonder meer een leereffect heeft. Het blijkt dat feedback zelfs een negatief effect kan hebben op leren als het niet op de juiste manier wordt vormgegeven (Boud & Malloy, 2013).

Over dit laatste fenomeen, dat feedback niet zonder meer een positief effect heeft bestaan verschillende publicaties (zie onder meer Glover & Brown, 2006; Price, Handley, Millar, & O'Donovan, 2010; Walker, 2009). Het tegenvallende resultaat van de feedback blijkt dikwijls te herleiden tot het niet naleven van de criteria voor goede feedback. Feedback staat dan bijvoorbeeld in het teken van legitimering van een cijfer en niet in het teken van ontwikkeling van de student (Butler, 1988; Glover & Brown, 2006). Waar feedback onderdeel uit zou moeten maken van een cyclisch proces wordt het vaak onderdeel gemaakt van de becijfering. Daarmee wordt de definitie van feedback niet waargemaakt. De definitie geeft immers aan dat het dient te gaan om informatie welke gebruikt wordt (*'which is used'*). Dit gegeven vraagt een heroverweging van de positie van feedback in het curriculum (zie ook Price et al., 2010).

Binnen het hoger onderwijs wordt een substantieel deel van de feedback schriftelijk gecommuniceerd. Gegeven het sterke effect dat formatieve toetsing kan hebben op het leren van de student (Hattie, 2009) is het zeer interessant om dit aspect verder te ontwikkelen binnen de lerarenopleiding. Docenten besteden al veel tijd en energie aan het formuleren van feedback, maar de vraag is of de

aanpak die nu gehanteerd wordt wel tot een leereffect leidt. Zo niet, welke aanbevelingen kunnen er gemaakt worden? Welke mogelijkheden zijn er om de huidige praktijk te verbeteren?

Het onderzoek dat in deze praktijkpublicatie besproken wordt heeft de vorm aangenomen van een actieonderzoek, waarbij er in dit hoofdstuk in beeld is gebracht welke aspecten van de praktijk voor verbetering vatbaar zijn (de verkenning) en waarbij er vervolgens een gerichte verbeteractie is ingezet, welke in het volgende hoofdstuk wordt besproken. Een volledig onderzoeksartikel is op te vragen bij de auteurs (Arts, Jaspers, & Joosten-ten Brinke, 2014).

2.3 Onderzoeksvragen

Bij de start van het onderzoek stond onderstaande onderzoeksvraag centraal.

In hoeverre voldoet de schriftelijke feedback van de docenten binnen de lerarenopleiding Biologie aan kwaliteitscriteria beschreven in de literatuur?

2.4 Methode

Context

Om zicht te krijgen op de kwaliteit van schriftelijke feedback binnen de opleiding Biologie van Fontys lerarenopleiding is de huidige praktijk geanalyseerd. Het onderzoek richt zich op feedback van docenten op het zogenaamde 'beslisdocument' van het afstudeeronderzoek

van de Bachelor opleiding. Dit beslisdocument wordt halverwege het schooljaar beoordeeld (Go of NoGo) om de haalbaarheid van het afstudeeronderzoek vast te stellen. De keuze om juist dit studieonderdeel te bestuderen op feedback, is ingegeven door het feit dat er over het algemeen veel feedback wordt gegeven aan de studenten en dat hier meerdere docenten bij betrokken zijn. Hoewel het niet een zuiver formatief moment is, gaat het wel om een document dat niet becijferd wordt. De hoeveelheid feedback en het gegeven dat deze van meerdere docenten afkomstig is maakt het aannemelijk dat de manier waarop feedback wordt gegeven binnen deze cursus min of meer illustratief is voor schriftelijke feedback in de gehele opleiding.

Deelnemers

Alle vierdejaars voltijd studenten die in schooljaar 2010 - 2011 deelnamen aan de cursus actieonderzoek zijn middels een vragenlijst bevraagd op de ervaringen (n=19). Daarnaast is de schriftelijke feedback van de drie docenten, die betrokken waren bij de cursus actieonderzoek geanalyseerd. Docent 1 is auteur van deze publicatie en was voor het derde jaar betrokken bij actieonderzoek, docent 2 was voor het eerst betrokken bij actieonderzoek, terwijl docent 3 voor het vijfde jaar de cursus uitvoerde.

Instrumenten

De perceptie van studenten is vastgesteld met behulp van een vragenlijst (opgesteld naar inzichten uit de literatuur, o.a. Glover & Brown, 2006; Hattie & Timperley, 2007).

De vragenlijst bestaat uit 12 open vragen gericht op kwaliteitscriteria waaraan schriftelijke feedback zou moeten voldoen, hoe studenten de feedback verwerken en welke behoeften studenten hebben met betrekking tot feedback.

Procedure

De vragenlijst is tijdens een klassikale bijeenkomst op papier ingevuld door de studenten, aan het einde van een reguliere les. Dit was een week nadat alle deelnemers feedback hadden ontvangen op het beslisdocument.

Data analyse

De antwoorden van de studenten op de vragenlijsten zijn kwalitatief geanalyseerd op inhoud en met elkaar vergeleken. Wanneer antwoorden van twee of meer studenten overeen kwamen, dan werden de gegevens meegenomen voor verdere analyse. Elke vraag uit de vragenlijst is gekoppeld aan succescriteria voor effectieve feedback zoals geformuleerd door Bruno en Santos (2010), Gibbs en Simpson (2004) en Nicol en Macfarlane-Dick (2006).

De schriftelijke feedback, bestaande uit ingevoegde opmerkingen van drie docenten is geanalyseerd en ingedeeld in categorieën, ingegeven door (i) het model van Hattie & Timperley (2007) en (ii) het model van Glover & Brown (2006). De opmerkingen (n=299) bij acht verschillende beslisdocumenten zijn hierbij gecategoriseerd. De eerste en een onafhankelijke onderzoeker hebben het eerste document (met 69

opmerkingen) onafhankelijk van elkaar geanalyseerd en vervolgens uitgewisseld. Daarbij werd een overeenstemming (kappa) van 0,193 bereikt. Vervolgens is er overleg geweest over de verschillende definities en is er een tweede document (met 46 opmerkingen) onafhankelijk van elkaar geanalyseerd. Nu werd er een overeenstemming (kappa) van 0,765 bereikt.

2.5 Resultaten

Kwaliteit van feedback: de perceptie van studenten

In Tabel 2.1 worden de voornaamste resultaten van de vragenlijst samengevat in de twee linker kolommen. In de derde kolom worden de resultaten gekoppeld aan kwaliteitscriteria uit de literatuur (Bruno & Santos, 2010; Gibbs & Simpson, 2004; Nicol & Macfarlane-Dick, 2006).

Tabel 2.1 Kwaliteit van de feedback volgens studenten (n=19).

Vraagstelling	Resultaat	Criterium (zie ook inleiding)
Geef kort aan wat de doelen en beoordelingscriteria van actieonderzoek zijn.	Studenten zijn erg algemeen in de formulering. Er worden geen concrete doelen of criteria benoemd.	Effectieve feedback: • Maakt de doelen van de opdracht duidelijk; • Maakt de beoordelingscriteria duidelijk; • Maakt het verwachte eindniveau duidelijk.
Wanneer is schriftelijke feedback waardevol voor jou (of voor jouw actieonderzoek)?	Er zijn geen opmerkingen gemaakt, die gekoppeld kunnen worden aan motivatie of zelfvertrouwen.	Effectieve feedback ondersteunt motivatie en zelfvertrouwen.
Wat zijn de sterkste punten van je actieonderzoek? Waar ben je goed in m.b.t. deze cursus en wat zou je nog kunnen verbeteren?	Studenten waren zeer divers in de beantwoording van deze vraag. Drie studenten geven geen sterke punten aan, maar ook geen verbeterpunten. Twee studenten geven enkel verbeterpunten aan.	Effectieve feedback helpt de student bij adequate zelf beoordeling.

Verwerking van feedback door studenten

De vragenlijst bevat ook enkele vragen met betrekking tot de verwerking van de ontvangen feedback. Eén daarvan is "Hoe verwerk je de geschreven feedback?". Dertien studenten geven aan dat de opmerkingen één voor één verwerkt worden. Vier studenten geven aan dat ze de feedback doorlezen en vervolgens een keuze maken met welke feedback ze aan de slag gaan. Eén student geeft aan de meest voorname opmerkingen uit te kiezen en dat deze verwerkt worden.

Bruikbaarheid van feedback voor studenten

De vragenlijst eindigt met de vraag "Wanneer is schriftelijke feedback waardevol voor jou (of voor jouw actieonderzoek)". De antwoorden van de studenten lopen voor deze vraag behoorlijk uiteen. Zeven studenten geven aan "wanneer ik er verder mee kan". Drie studenten geven aan "wanneer het concreet is". Twee studenten antwoordden "Wanneer het je actieonderzoek naar een hoger plan tilt".

Analyse van feedback

De acht beslisdocumenten die geanalyseerd zijn bevatten in totaal 299 opmerkingen (*gemiddeld* = 37; *sd* = 19).

Diepte van feedback

De opmerkingen werden als eerste geanalyseerd en gecategoriseerd op diepte. De resultaten zijn in Tabel 2.2 weergegeven. Het type feedback dat het meest gegeven wordt, is een indicatie (diepte 1) of iets goed is of niet, gevolgd door verbeteringen (diepte 2). Slechts 12 % van

de opmerkingen bevatten ook een uitleg of toelichting (diepte 3).

Tabel 2.2 Verdeling van opmerkingen over drie diepten van feedback (n=299)

Diepte	n	%
Indicatie	201	67
Correctie	63	21
Indicatie / Correctie + Uitleg/toelichting	35	12

Type feedback

Opmerkingen zijn ook geanalyseerd op niveau (taak, proces, zelfregulatie of persoon). De resultaten van de analyse zijn in Tabel 2.3 weergegeven. De meeste opmerkingen bevatten feedback op taakniveau, de opmerkingen met betrekking tot proces en zelfregulatie vormen samen ongeveer de helft van het totaal. Opmerkingen gericht op de persoon komen nauwelijks voor.

Tabel 2.3 Verdeling van opmerkingen over verschillende feedback niveaus (n = 299)

Niveau	n	%
Taak	144	48,1
Proces	79	26,4
Zelfregulatie	72	24,1
Persoon	4	1,3

Feedback, feed up en feed forward

Tot slot zijn de opmerkingen geanalyseerd op de aspecten feedback, feed up en feed forward. Hieruit werd duidelijk dat feed up en feed forward schaars gehanteerd werden

(zie Tabel 2.4). Opmerkingen bevatten vrijwel zonder uitzondering informatie met betrekking tot de vraag “hoe doe ik het tot nu toe?”. Informatie die als feed up geclassificeerd kan worden werd sporadisch aangetroffen, terwijl feed forward nergens expliciet naar voren kwam.

Tabel 2.4 *Verdeling van opmerkingen over verschillende feedback categorieën (n=299)*

Categorie	%
Feed back	88
Feed up	12
Feed forward	0

2.6 Discussie & Conclusies

Kwaliteit van feedback

Op basis van de antwoorden op de vragenlijst lijken studenten over het algemeen tevreden over de kwaliteit van de schriftelijke feedback. De resultaten maken echter wel duidelijk dat er ruimte voor verbetering is. De uitkomsten van deze casestudy sluiten in grote lijnen aan bij uitkomsten van andere onderzoeken waarbij feedback geanalyseerd is (zie onder meer Buhagiar, 2012; Glover & Brown, 2006; Walker, 2009; Weaver, 2006) en biedt een uitbreiding van deze onderzoeken. Zo is het (voor zover ons bekend) de eerste studie waarbij het model van Hattie en Timperley (2007) ingezet is voor categorisatie van feedback. De feedback praktijk voldoet volgens de studenten aan de volgende aspecten voor effectieve feedback: effectieve feedback wordt tijdig gegeven (binnen 2 weken) en is helder geformuleerd (resultaten

niet weergegeven). Aan andere succescriteria wordt niet, of slechts gedeeltelijk voldaan.

Studenten waren niet in staat om doelen en beoordelingscriteria te beschrijven. Wanneer doelen en beoordelingscriteria niet eigen gemaakt zijn, dan is het onwaarschijnlijk dat studenten een duidelijk beeld hebben van het verwachte eindniveau. Dit zegt natuurlijk iets over de instructie tijdens bijeenkomsten, maar daarnaast mag er voorzichtig geconcludeerd worden dat de schriftelijke feedback niet refereert aan criteria en doelen en dat de feedback de studenten daardoor weinig houvast biedt met betrekking tot het verwachte eindniveau. Opvallend genoeg is er een student die op de vragenlijst aangeeft dat de feedback niet volledig werd begrepen doordat het onduidelijk is wat het eindniveau is. Een zwakke correlatie tussen feedback en beoordelingscriteria komt ook in andere onderzoeken terug (Ferguson, 2009; Weaver, 2006).

Studenten geven aan dat ze bij het aanpassen van het beslisdocument opmerkingen één voor één verwerken. Daarmee lijkt het zo te zijn dat studenten in deze casestudy de feedback met name op taakniveau gebruiken. Dit zou indirect een gevolg kunnen zijn van de benadering van de docenten, waarbij (voor dit curriculumonderdeel) met name het verbeteren van de prestatie centraal staat en niet zozeer het verbeteren van strategieën, vaardigheden en ontwikkeling van de student. Het voorgaande is ook beschreven door Orsmond en Merry (2011).

Opvallend genoeg geven studenten aan dat feedback bruikbaar is wanneer het helpt om hun product te verbeteren. Dit impliceert dat een effect op leren over het hoofd gezien wordt en dat het voornaamste doel bestaat uit het halen van de toets. Om tot een verandering van deze oppervlakkige benadering te komen en om studenten te stimuleren tot duurzaam leren, lijkt er een verschuiving van *assessment of learning* naar *assessment for learning* nodig. In het huidige curriculum lijkt er niet zo veel urgentie om te leren van de ontvangen feedback. Wanneer het curriculum studenten niet voorbereidt op het adequaat gebruik van feedback, dan zou het niet al te verrassend moeten zijn dat studenten een oppervlakkige benadering kiezen. Een meer centrale rol voor feedback en het daadwerkelijk gebruik daarvan in het curriculum zou wel eens nodig kunnen zijn om dit om te buigen (Black & Wiliam, 1998; Hernández, 2012; Mutch, 2003; Orrell, 2006). Daarnaast geven studenten aan dat ze feedback nuttig vinden als het ze in staat stelt om verder te gaan. Dit sluit aan bij de manier waarop ze feedback verwerken. Het lijkt te suggereren dat feedback op taakniveau geprefereerd wordt boven feedback op proces en zelfregulatie. Feedback op zelfregulatie vraagt vaak kritische reflectie en is minder rechtlijnig dan feedback op de taak. Een voorkeur voor feedback op concrete en eenvoudig aan te passen onderwerpen boven meer algemene feedback over meer abstracte aspecten is ook beschreven door Buhagian (2012), Cohen (1987) en Hernández (2012). In het onderzoek van Ferguson (2009), echter, geven studenten juist aan dat zij een voorkeur hebben voor algemene zaken. Het is voor de

huidige casestudy van belang om te benadrukken dat de uitkomst van deze vraag waarschijnlijk beïnvloed is door het gegeven dat het vierdejaarsstudenten betreft. Het ligt vrij voor de hand dat zij voornamelijk willen afstuderen. In relatie tot de literatuur (Ferguson, 2009; Glover & Brown, 2006), lijken de opmerkingen van de studenten aan te sluiten bij het gegeven dat zij zicht willen krijgen op *hoe* ze “*the gap*” kunnen slechten. Studenten schrijven niets in relatie tot motivatie of zelfvertrouwen. Dit in tegenstelling tot de studie van Glover en Brown (2006) en Weaver (2006). Dit kan mogelijk verklaard worden door de context van een kleine groep studenten (n=19) en de regelmatige bijeenkomsten met de betrokken docenten.

Wanneer er gekeken wordt naar de analyse van de opmerkingen dan wordt duidelijk dat de docenten voornamelijk gebruik maken van indicaties en correcties als feedback. Uitleg is daarbij schaars. Dit sluit aan bij andere publicaties (Buhagiar, 2012; Glover & Brown, 2006; Hernández, 2012; Walker, 2009) en dat zou kunnen betekenen dat het een vrij algemeen fenomeen is in hoger onderwijs. Zoals in de inleiding al aangegeven blijft hiermee de gap bestaan. Het zou kunnen zijn dat uitleg achterwege gelaten wordt uit tijdsoverwegingen, zoals ook al gesuggereerd in de literatuur (Hernández, 2012; Orrell, 2006; Weaver, 2006). Het zou ook kunnen betekenen dat docenten de urgentie van uitleg of toelichting niet inzien (Gibbs & Simpson, 2004).

In deze casestudy is er gebruik gemaakt van het theoretische model van Hattie en Timperley (2007) om

de feedback te categoriseren. Hoewel de beschrijving in dit artikel behoorlijk duidelijk is, bleek het vrij lastig om het daadwerkelijk in te zetten voor het rubriceren van opmerkingen. Classificatie van opmerkingen met betrekking tot taak en persoon was vrij eenvoudig. Het onderverdelen van opmerkingen in proces of zelfregulatie, was veel lastiger. Een opmerking als “betere focus op insteek: leerlingen actief laten leren” werd ingedeeld als feedback op procesniveau door een onderzoeker, aangezien het te maken heeft met de gekozen aanpak of strategie van de student, terwijl de andere onderzoeker koos voor een indeling bij zelfregulatie, aangezien het van de student vraagt om nog eens kritisch naar de inhoud van het verslag te kijken. De initiële mate van overeenstemming bij het rubriceren van 69 opmerkingen was dan ook slechts 0,193. Na het uitvoerig bespreken van definities was er bij de indeling van de opmerkingen bij een tweede verslag weliswaar meer overeenkomst (0,765) tussen de eerste en laatste onderzoeker, maar indeling van opmerkingen leverde ook daarna nog discussie met andere onderzoekers. Deze meningsverschillen kunnen gedeeltelijk teruggevoerd worden op de niveauduiding in het artikel van Hattie en Timperley (2007), waarbij feedback op procesniveau gekoppeld wordt aan de term ‘strategie’, terwijl bijvoorbeeld Nicol en Macfarlane-Dick (2006) en Butler en Winne (1995) de term ‘strategie’ in de context plaatsen van zelfregulatie. Dit roept wel enkele vragen op met betrekking tot de bruikbaarheid van (een klein gedeelte van) het model van Hattie en Timperley (2007). Misschien is het wel zo dat het onderscheid tussen feedback op proces en zelfregulatie zo moeilijk te maken is

doordat het een kunstmatig onderscheid is.

Uit literatuur komt naar voren dat feedback op het niveau van proces en zelfregulatie het sterkste leereffect heeft (Hattie & Timperley, 2007). Dit soort opmerkingen omvatten ongeveer de helft van de geanalyseerde opmerkingen. Verbetering van de schriftelijke feedback zou misschien mogelijk zijn als docenten beter op de hoogte zijn van de verschillen tussen de verschillende niveaus en de effecten daarvan.

De geanalyseerde feedback werd gegeven op een beslisdocument halverwege een schooljaar. Opvallend genoeg werd er geen feed forward gevonden en vrijwel geen feed up (12 %). Wanneer er geen feed forward gegeven wordt op een tussenversie van het eindverslag, waar de bruikbaarheid voor de student in de nabije toekomst ligt, wat zegt dat dan over feed forward in een breder perspectief? Het lijkt er op dat de docenten in deze casestudy min of meer onbekend zijn met een toekomstige dimensie van de feedback. Ook dit is in een eerdere studie naar voren gekomen (Buhagiar, 2012; Hernández, 2012). Het geringe gebruik van feed up sluit aan bij de antwoorden op de vragenlijst. De betrokken docenten geven vrijwel (direct) geen antwoord op de vraag ‘waar werk ik naartoe?’. Dan is het ook niet vreemd dat studenten het lastig vinden om doelen en beoordelingscriteria te formuleren.

In hoeverre voldoet de schriftelijke feedback van de docenten binnen de lerarenopleiding Biologie aan kwaliteitscriteria beschreven in de literatuur?

Hoewel er aan verschillende kwaliteitscriteria voor effectieve feedback voldaan wordt, legt deze casestudy enkele aspecten bloot die verbeterd kunnen worden. De belangrijkste aspecten die het leereffect van de feedback (waarschijnlijk) zal laten toenemen zijn:

- expliciete koppeling van feedback aan doelen, beoordelingscriteria en verwacht eindniveau;
- uitleg/toelichting (diepte 3) te geven als aanvulling op indicaties (diepte 1) en correcties (diepte 2);
- meer focus op feedback op het niveau van proces en zelfregulatie, in plaats van op taakniveau;
- meer gebruik van feed up en feed forward.

Voor zover ons bekend, is dit de eerste studie waarbij de classificatie volgens het model van Hattie en Timperley (2007) wordt toegepast. Hopelijk zullen er meerdere onderzoeken volgen waarbij deze classificatie wordt toegepast en leidt dat een verdere uitwerking van het model. De hier gepresenteerde casestudy combineert een analyse van feedback en de perceptie van de studenten, wat vrij zeldzaam is wanneer er naar andere onderzoeken gekeken wordt. Daarmee vormt het zowel een aanvulling als een uitbreiding op publicaties als die van Hernández (2012), Glover en Brown (2006), Walker (2009) en Weaver (2006). De indruk is dat de gevonden resultaten exemplarisch zijn voor het (hoger) onderwijs en om tot meer efficiënte en effectieve feedback te komen stellen we voor om gebruik te maken van een vormgever.

Literatuur (Newton et al., 2012) laat zien dat het gebruik van een (digitale) vormgever (een overzicht met hulpvragen voor de docent) bij kan dragen aan het verbeteren van de schriftelijke feedback. Hiervoor verwijzen we naar het volgende hoofdstuk.

Vier jaren kenniskring: een terugblik van Jorik Arts

Wat een ontzettend leerzame, inspirerende en leuke ervaring om deel uit te maken van de kenniskring van het lectoraat 'Eigentijds toetsen en beoordelen'! Dat is eigenlijk de beste samenvatting die ik hier kan geven. Kennis delen en opdoen samen met de andere kenniskringleden, samenwerken met Mieke, literatuuronderzoek, afstand nemen van de dagelijkse praktijk, praktijkonderzoek opzetten en uitvoeren, het is een heel scala aan activiteiten dat gekoppeld is aan de kenniskring, voorgezeten door onze bevoegde lector Desiree Joosten - ten Brinke. Los van alles wat ik geleerd heb omtrent assessment for learning en feedback in het bijzonder, denk ik dat het zelf doen van praktijkonderzoek ook positief bijdraagt aan de manier waarop ik praktijkonderzoek van studenten begeleidt. Het zelf ervaren van allerlei problemen bij de uitvoering van praktijkonderzoek vergroot natuurlijk de herkenning van problemen die studenten ervaren.

Vier jaren kenniskring: een terugblik van Mieke Jaspers

In de eerste plaats was het voor mij een zeer waardevolle ervaring (zo niet luxe) om vier jaren met één onderwerp de diepte in te kunnen gaan. Een en ander daagt je uit om uitgebreid te zoeken naar relevante literatuur, te experimenteren aan de hand van de bevindingen uit reeds gedaan onderzoek en tegelijkertijd te professionaliseren op het gebied van de effectiviteit en de efficiëntie van schriftelijke feedback. Een ander belangrijk onderdeel van vier jaren actieve deelname in de kenniskring was in eerste plaats de constructieve samenwerking met Jorik Arts.

Voor mij werkt het om samen aan de slag te gaan, elkaar aan te vullen en te enthousiasmeren. Daarnaast hebben de frequente kenniskringbijeenkomsten, vol energie en deskundigheid geleid door de lector Desiree Joosten - ten Brinke, aangezet tot kennisdeling, kritisch kijken en projectresultaat opleveren. En daar houd ik ook van.

Terugblikkend op vier waardevolle en leerzame jaren in de kenniskring kan ik aangeven dat ik extra kennisbagage heb ontwikkeld op het gebied van toetsing in het bijzonder assessment for learning, hetgeen meerwaarde heeft in mijn functie als onderwijskundig adviseur.

Hoofdstuk 3

De meerwaarde van een vormgever

Jorik Arts en Mieke Jaspers

3.1 Samenvatting

Feedback heeft een positief effect op het leren. Voorwaardelijk is echter dat de feedback voldoet aan enkele succescriteria. Afgaand op het aantal publicaties waarin problemen rondom het formuleren van effectieve feedback wordt beschreven is het lang niet makkelijk om de handvatten uit de literatuur in de praktijk te brengen. In deze studie wordt een hulpmiddel (een vormgever) beschreven die is ingezet voor verbetering van de kwaliteit van schriftelijke feedback. Analyse van de schriftelijke feedback laat zien dat het gebruik van de vormgever leidt tot meer aandacht voor feed up en feed forward en dat er meer feedback op procesniveau wordt geformuleerd.

3.2 Inleiding

Feedback kan een sterk positief effect hebben op het leren, mits deze op de juiste wijze wordt gegeven (zie o.a. Hattie & Timperley, 2007). Er bestaat echter nogal wat literatuur waaruit blijkt dat het op de juiste wijze formuleren van feedback niet zondermeer goed verloopt in het (hoger) onderwijs (Boud & Malloy, 2013; Glover & Brown, 2006; Price, Handley, Millar, & O'Donovan, 2010; Walker, 2009). Als mogelijke oorzaak voor dit gegeven wordt verwezen naar gebrek aan tijd, maar ook naar een gebrek aan kennis omtrent effectieve feedback bij docenten (Orrell, 2006; Price, et al., 2010; Walker, 2009). Deze assumptie wordt versterkt door het gegeven dat veel aspecten die van doen hebben met (formatieve) toetsing in het hoger onderwijs al doende eigen gemaakt zijn, zonder dat daar expliciete scholing aan vooraf is gegaan (Perera, et al., 2008; Price, et al., 2010). Mogelijk dat de problematiek rondom het formuleren van effectieve feedback nog versterkt wordt doordat docenten in hun handelen niet altijd congruent blijken met de opvattingen die zij communiceren (Adcroft, 2011; Lee, 2009; Perera, et

al., 2008). Lee (2009) beschrijft tien voorbeelden waarbij de daadwerkelijke feedbackpraktijk verschilt van de overtuigingen van docenten die in interviews naar voren gebracht worden. Zo geven docenten bijvoorbeeld aan dat ze studenten aan willen moedigen om zelf fouten te leren herkennen, terwijl ze in de praktijk uitvoerig verbeteringen aanbrengen in studentwerk of fouten markeren. Daarnaast geven docenten cijfers voor producten, ondanks de overtuiging dat dit de aandacht afleidt van de daadwerkelijke feedback. En zo beschrijft Lee (2009) nog acht andere voorbeelden. De weerbarstigheid van de praktijk?

Bij het geven van schriftelijke feedback lijken er in hoofdlijnen vier aspecten te zijn, waarop het mis lijkt te lopen:

- Feedback wordt vaak gegeven in de vorm van opmerkingen of correcties zonder toelichting (Glover & Brown, 2006; Lee, 2009; Walker, 2009);
- Feedback wordt gecombineerd met een cijfer (Butler, 1988; Lee, 2009);
- Feedback richt zich voornamelijk op de uitgevoerde taak, die vaak niet meer in dezelfde vorm terug komt in het curriculum, en is in veel mindere mate gericht op toekomstige taken of de ontwikkeling van de ontvanger (Carless, 2006; Glover & Brown, 2006; Hyatt, 2005; Lee, 2009; Orsmond & Merry, 2011; Perera, et al., 2008);
- Feedback richt zich met name op zwakten en veel minder op sterktes (Glover & Brown, 2006; Lee, 2009).

Het in praktijk brengen van de aandachtspunten uit de literatuur voor goede feedback is blijkbaar lastig. Of dat nu komt door gebrek aan scholing of door gebrek aan tijd is niet geheel duidelijk.

Gekeken naar de positie die formatief bedoelde feedback, als onderdeel of onderbouwing van een summatieve beoordeling, in veel gevallen inneemt in het curriculum is het misschien niet zo raar dat de effecten niet altijd zijn zoals docenten dat zouden willen. Zeker in het hoger onderwijs worden studenten geconfronteerd met sterk modulair onderwijs waarbij ze te maken krijgen met verschillende vakken die verzorgd worden door verschillende docenten met allemaal eigen wensen en verwachtingen (Price, et al., 2010). Dit maakt ontwikkelingsgerichte feedback lastig, aangezien het voor de betrokken docenten niet zondermeer duidelijk is welke opdrachten studenten in het nabije verleden hebben gehad of welke zij in de nabije toekomst gaan krijgen (Price, et al., 2010). Het voorgaande vraagt nadrukkelijke afstemming op het niveau van het totale toetsprogramma. Misschien is het in deze context te overwegen om in sommige gevallen feedback (als onderbouwing bij een summatieve beoordeling) te minimaliseren, bijvoorbeeld door het gebruik van een rubric, zodat er tijd vrij komt om formatieve feedback te genereren die ook daadwerkelijk gekwalificeerd kan worden als feedback (Glover & Brown, 2006).

Hoewel de aangehaalde onderzoeken zich met name richten op het hoger onderwijs, ligt het voor de hand om

te veronderstellen dat het voortgezet onderwijs niet veel af zal wijken (Lee, 2008). Binnen een lerarenopleiding is het daarmee van belang dat een goed voorbeeld wordt neergezet voor aanstaande docenten, zodat zij dat door kunnen zetten naar de eigen lespraktijk (Buhagiar, 2012)

Welke mogelijkheden zijn er om tot een verandering in feedbackpraktijk te komen? Uit literatuuronderzoek komt naar voren dat een vormgever, dit is een voorblad met enkele hulpvragen, docenten kan helpen bij het formuleren van feedback (Newton, et al., 2012). In het onderzoek van Newton et al. (2012) is met name ingezet op wat Hattie & Timperley (2007) aanduiden als feed up, met per beoordelingscriterium twee hulpvragen voor de docent: wat is er goed?, welke aanpassingen zouden leiden tot een betere beoordeling? Het gebruik van een vormgever resulteerde in deze studie in (i) meer feedback en in (ii) meer uitleg bij de feedback (Newton, et al., 2012).

In het vooronderzoek (zie hoofdstuk 2) is voor de context van deze case-study vastgesteld dat verbetering van de feedbackpraktijk mogelijk bereikt kan worden wanneer er meer aandacht komt voor:

- diepte 3 / uitleg bij de feedback (Glover & Brown, 2006);
- feed up en feed forward (Hattie & Timperley, 2007);
- feedback op het niveau van proces en zelf regulatie (Hattie & Timperley, 2007).

Glover en Brown (2006) delen feedback op in diepte 1,2 en 3. Onder diepte 1 wordt verstaan dat feedback bestaat uit een opmerking of indicatie (bijv. “prima”), bij diepte 2 is er sprake van een correctie of verbetering (bijv. “deze tekst verplaatsen naar het voorwoord”), bij diepte 3 is er sprake van een uitleg bij een opmerking of een verbetering.

In het model van Hattie en Timperley (2007) wordt aangegeven dat effectieve feedback helderheid verschaft over het werk dat verzet is (feedback) en over doelen, criteria en verwachtingsniveau. Dat laatste noemen zij feed up. Daarnaast biedt effectieve feedback handvatten met betrekking tot toekomstige opdrachten (feed forward). Voor elk type feedback (feedback, feed up en feed forward) geldt dat er nog een verdere onderverdeling mogelijk is in feedbackniveau. De vier niveaus die daarbij geduid worden zijn taak (bijv. is bepaalde informatie correct / incorrect), proces (bijv. hoe is de werkwijze / strategie), zelfregulatie (bijv. wat zijn sterktes / zwaktes) en persoon (bijv. je bent een creatieve student). Met name feedback op het niveau van proces en zelfregulatie blijken een sterk effect op leren te hebben (Hattie & Timperley, 2007).

Deze aspecten zijn dan ook allemaal meegenomen in de ontwikkeling van de vormgever. In het hier beschreven onderzoek worden de (voorlopige) resultaten van het gebruik van de vormgever (zie Figuur 3.1) weergegeven.

Vormgever effectief en efficiënt feedback geven

Algemene indruk: benoem opvallende zaken en probeer motiverend te zijn in het taalgebruik.

Je bent duidelijk op de goede weg. Grote delen van je verslag zitten al goed in elkaar, zoals methode van onderzoek voor de verkenning en de presentatie van de gegevens van de IIn. Wanneer je de huidige opzet aanvult met wat informatie dan zit je op koers voor een GO.

Feedback:

Welke onderdelen zijn goed of minder goed en waarom (taak, proces)?

De beschrijving van de methode van onderzoek en de presentatie van je onderzoeksgegevens (interview + enquêtes) is heel duidelijk verwoord (beoordelingsformulier: informatieverwerking). Bij de verbeteracties geef je al heel duidelijk aan wat je waarom verwacht aan resultaten (synthese en onderbouwing)

Er mist nog wel wat aan de literatuurverkenning. Deelvraag 6 wordt nog niet beantwoord. Daarnaast zou de rode draad wat meer nadruk mogen hebben bij de bespreking van literatuur. Je kan hieraan werken door iets vaker een tussenconclusie te geven of te koppelen aan jouw situatie. (leesbaarheid en navolgbaarheid & informatieverzameling)

Feed Up: koppelen aan doelen / criteria (wat zijn essentiële aanpassingen)

Leg uit hoe je de verbeteracties nu precies gaat uitvoeren in je lessen. Dat kan bijvoorbeeld door te verwijzen naar lesvoorbereidingsformulieren. (*beoordelingsformulier leesbaarheid en navolgbaarheid*).

Maak de literatuurbespreking duidelijker voor de lezer en werk deelvraag 6 uit. Neem nog een aanvullende deelvraag op in je verslag: "hoe groot is het probleem?" (*verwerking van literatuur*).

Feed Forward (koppelen aan het leren van de student)

Probeer nog duidelijker (concreter) de dwarsverbanden in je verslag aan te geven. Verwijs dus naar de antwoorden op deelvragen bij de verbeteracties, verwijs bij de ervaring van leerlingen naar de uitkomst van literatuurvragen, etc. Dan wordt het voor de lezer (nog) duidelijker dat je de stof eigen hebt gemaakt. De uitwerking van methode van onderzoek bij de verkenning en de presentatie van gegevens bij IIn en collega's is erg duidelijk. Neem dat mee voor de beschrijving van de verbeteracties en de methode van onderzoek naar het effect van de verbeteracties.

Feed Forward

Op welke gemaakte keuzes moet de student nog reflecteren (zelf regulatie)?

Onderstaande opmerkingen hoef je niet aan te passen in je verslag (het mag wel), maar deze vragen zouden terug kunnen komen bij je eindpresentatie:

Wat zegt de verhoging van 0,8 punt eigenlijk? Wat is nu het onderliggende doel wat je wilt bereiken? Waarom vind jij de verhoging van het cijfer belangrijk? (om over na te denken voor bij je presentatie)

Figuur 3.1 De vormgever (ingevuld door een van de docenten).

3.3 Onderzoeksvragen

In dit onderzoek staat de volgende onderzoeksvraag centraal:

In hoeverre leidt het gebruik van een vormgever voor feedback binnen de lerarenopleiding Biologie tot kwalitatief goede schriftelijke feedback?

Hierbij wordt onder kwalitatief goede feedback verstaan dat de feedback bestaat uit feedback, feed up en feed forward, dat het voldoende gericht is op proces en zelfregulatie (Hattie & Timperley, 2007) en dat de feedback uitleg bevat (Glover & Brown, 2006). Daarnaast zijn de aanbevelingen van Nicol en Macfarlane-Dick (2006) en Bruno en Santos (2010) meegenomen.

3.4 Methode

Context

Om zicht te krijgen op de kwaliteit van schriftelijke feedback binnen de opleiding Biologie van de Fontys lerarenopleiding, is er gekeken naar feedback van docenten op het zogenaamde ‘beslisdocument’ van het afstudeeronderzoek van de Bachelor opleiding. Dit beslisdocument wordt halverwege het schooljaar beoordeeld (Go of NoGo) om de haalbaarheid van het afstudeeronderzoek, een actieonderzoek, vast te stellen.

Deelnemers

De schriftelijke feedback van de vier docenten die betrokken waren bij de cursus actieonderzoek is

geanalyseerd. Docent 1 is auteur van deze publicatie en was voor het vijfde jaar betrokken bij actieonderzoek, docent 2 en docent 3 waren voor het derde jaar betrokken bij actieonderzoek, terwijl docent 4 voor het tweede jaar betrokken was. In totaal hebben 34 studenten deze cursus gevolgd (schooljaar 2012 - 2013).

Procedure

De beslisdocumenten werden door de studenten via email ingeleverd bij docent 1. Daarna werden de eerste twaalf documenten op volgorde van binnenkomst verdeeld over de docenten. Van de twaalf documenten zijn er drie willekeurig gekozen verslagen enkel met opmerkingen van feedback voorzien. De overige documenten zijn door de docenten voorzien van feedback op de ‘traditionele’ manier met (digitaal) ingevoegde opmerkingen bij de tekst en met feedback via de vormgever. Voorafgaand aan het formuleren van feedback is (i) het model van Hattie en Timperley (2007) besproken met alle betrokken docenten, is er (ii) stilgestaan bij enkele probleempunten rondom feedback aan de hand van onder andere het artikel van Walker (2009) en (iii) is de bedoeling (effectieve & efficiënte feedback) van het gebruik van de vormgever besproken met de betrokken docenten. De feedback werd digitaal gegeven door de docenten en na afronding verzonden naar de studenten en naar docent 1 voor analysedoeleinden.

Data analyse

De schriftelijke feedback, bestaande uit ingevoegde opmerkingen van vier docenten is kwalitatief geanalyseerd

op basis van een codering ingegeven door (i) het model van Hattie & Timperley (2007) en (ii) het model van Glover & Brown (2006). De opmerkingen ($n=554$) bij twaalf verschillende beslisdocumenten zijn hierbij gecategoriseerd. Om te bepalen of het gebruik van de vormgever leidt tot een andere aanpak bij het formuleren van feedback bij de verslagen zijn de opmerkingen geanalyseerd op (i) diepte, (ii) feedbackniveau en (iii) type feedback (feedback, feed up en feed forward). Tot slot zijn de opmerkingen op de vormgever ($n=173$) op een vergelijkbare manier geanalyseerd als de opmerkingen bij de verslagen zelf en is er gekeken naar het aantal opmerkingen per rubriek van de vormgever (de vormgever is ingedeeld in verschillende rubrieken: algemene indruk, feedback, feed up en feed forward). Daarbij is nagegaan of de teksten in de respectievelijke rubrieken ook daadwerkelijk daar thuis horen. Zijn de opmerkingen herkenbaar als feedback, feed up of feed forward? Hierbij zijn alle opmerkingen op formulering bekeken (validiteit van de vormgever). Om te bepalen of er een significant verschil bestaat tussen de onderverdeling van opmerkingen op de vormgever en de opmerkingen ingevoegd bij de verslagen is er een t-toets uitgevoerd. Omdat de hoeveelheid opmerkingen ingevoegd bij verslagen veel groter is dan de hoeveelheid opmerkingen op de vormgevers is hierbij gewerkt met de percentuele verdeling van opmerkingen over verschillende categorieën. De analyses zijn uitgevoerd door de eerste auteur van deze publicatie.

3.5 Resultaten

Bij het schrijven van deze praktijkpublicatie zijn nog niet alle resultaten geanalyseerd. Het betreft hier dus voorlopige resultaten.

Analyse van feedback: opmerkingen ingevoegd in het verslag

De twaalf beslisdocumenten die geanalyseerd zijn bevatten in totaal 554 opmerkingen (*gemiddeld* = 46; *sd* = 11,0). Drie verslagen zijn op de 'traditionele' wijze van feedback voorzien: zonder vormgever en met enkel ingevoegde opmerkingen. Deze verslagen bevatten in totaal 141 opmerkingen (*gemiddeld* = 47, *sd* = 11,0). Negen verslagen zijn van feedback voorzien met ingevoegde opmerkingen en daarnaast met behulp van de vormgever. Deze verslagen bevatten in totaal 413 opmerkingen (*gemiddeld* = 46, *sd* = 9,4).

Diepte van feedback

De opmerkingen werden als eerste geanalyseerd en gecategoriseerd op diepte. De resultaten zijn in Tabel 3.1 weergegeven. Het type feedback dat het meest gegeven wordt, is een indicatie of iets goed is of niet (diepte 1), gevolgd door verbeteringen (diepte 2). 12 % van de opmerkingen bevatten ook een uitleg of toelichting (diepte 3).

Tabel 3.1 Verdeling van opmerkingen over drie diepten van feedback (n=554).

Diepte	Zonder vormgever			
	n	%	gem	sd
Indicatie	95	67	31,7	10,8
Correctie	25	18	8,3	7,9
Indicatie/Correctie + Uitleg/toelichting	21	15	7,0	3,6

Diepte	Met vormgever			
	n	%	gem	sd
Indicatie	286	69	31,8	11,6
Correctie	79	19	10,0	7,0
Indicatie/Correctie + Uitleg/toelichting	48	12	5,3	3,5

Diepte	Totaal	
	n	%
Indicatie	381	69
Correctie	104	19
Indicatie/Correctie + Uitleg/toelichting	69	12

Niveau feedback

Opmerkingen zijn ook geanalyseerd op niveau (taak, proces, zelfregulatie of persoon). De resultaten van de analyse zijn in Tabel 3.2 weergegeven. De meeste opmerkingen bevatten feedback op taakniveau (43,5 %), de opmerkingen met betrekking tot proces (33%) en zelfregulatie (23,5%) vormen samen ongeveer de helft van het totaal. Opmerkingen gericht op de persoon komen niet voor.

Tabel 3.2 Verdeling van opmerkingen over verschillende feedbackniveaus (n = 554).

Niveau	Zonder vormgever			
	n	%	gem	sd
Taak	67	47	22,3	6,9
Proces	39	28	13,0	5,7
Zelfregulatie	35	25	11,7	5,4
Persoon	0	0	0	0

Niveau	Met vormgever			
	n	%	gem	sd
Taak	174	42	19,3	8,2
Proces	144	35	16,0	5,2
Zelfregulatie	95	23	10,6	5,2
Persoon	0	0	0	0

Niveau	Totaal	
	n	%
Taak	241	43,5
Proces	183	33
Zelfregulatie	130	23,5
Persoon	0	0

Feedback, feed up en feed forward

Tot slot zijn de opmerkingen geanalyseerd op de aspecten feedback, feed up en feed forward. Feed up en feed forward worden nauwelijks gebruikt (zie Tabel 3.3).

Tabel 3.3 Verdeling van opmerkingen over verschillende feedback categorieën (n=554).

Categorie	Zonder vormgever			
	n	%	gem	sd
Feed back	130	92	43,8	11,1
Feed up	8	6	3,0	1,9
Feed forward	3	2	1,3	0,8

Categorie	Met vormgever			
	n	%	gem	sd
Feed back	410	99	45,6	11,0
Feed up	0	0	0	0
Feed forward	3	1	0,3	0,5

Categorie	Totaal	
	n	%
Feed back	540	97,5
Feed up	8	1,4
Feed forward	6	1,1

Analyse van feedback op de vormgever

De hulpvragen op de vormgever zijn gesorteerd naar verschillende feedbackniveaus en op feedback, feed up en feed forward.

De feedback op de negen ingevulde vormgevers bij de beslisdocumenten bevatten in totaal 173 opmerkingen (*gemiddeld* = 19; *sd* = 7).

De vormgever is ingedeeld in verschillende rubrieken: algemene indruk, feedback, feed up (essentiële aanpassingen) en feed forward. Wanneer de eerste rubriek (algemene indruk) buiten beschouwing gelaten wordt, dan blijven er in totaal 142 opmerkingen over. Een kleine

meerderheid van de opmerkingen worden geplaatst in de rubriek feedback (60 opmerkingen; 42% van de opmerkingen). Er staan 36 opmerkingen in de rubriek feed up en 46 in de rubriek feed forward (zie Tabel 3.4).

Tabel 3.4 Aantal opmerkingen (n=142) in de verschillende rubrieken (feedback, feed up en feed forward) op de vormgever.

Rubriek	n	%	gem	sd
Feed back	60	42	6,7	3,9
Feed up	36	25	4,0	2,7
Feed forward	46	32	5,1	2,8

Diepte van feedback

De opmerkingen zijn ook geanalyseerd en gecategoriseerd op diepte. De vormgevers bevatten voornamelijk (81%) opmerkingen die als diepte 1 (indicatie) gecategoriseerd kunnen worden. De resultaten zijn in Tabel 3.5 weergegeven.

Tabel 3.5 Verdeling van opmerkingen op de vormgever over drie diepten van feedback (n=173)

Rubriek	n	%	gem	sd
Indicatie	140	81	15,6	7,2
Correctie	3	2	0,6	0,7
Indicatie / Correctie + Uitleg/ toelichting	30	17	3,3	2,1

Niveau feedback

Anders dan de opmerkingen die bij het verslag geplaatst zijn, vormt feedback op taakniveau niet langer de grootste categorie (36% van de opmerkingen). De meeste opmerkingen (45%) hebben betrekking op proces (aanpak/strategie). Opmerkingen op het niveau van zelfregulatie vormen 18% van het totaal. Opmerkingen op het niveau van persoon komen zeer zelden voor (1% van de opmerkingen). De resultaten van de analyse zijn in Tabel 3.6 weergegeven.

Tabel 3.6 Verdeling van opmerkingen op de vormgever over verschillende feedback niveaus (n = 173)

Rubriek	n	%	gem	sd
Taak	63	36	7,0	3,7
Proces	78	45	8,7	2,6
Zelfregulatie	31	18	3,4	2,7
Persoon	1	1	0,1	0,3

De verdeling van opmerkingen in verschillende feedbackniveaus (taak, proces, zelfregulatie en persoon) lijkt anders te zijn dan die van de opmerkingen ingevoegd bij de verslagen. Alleen voor de feedback op procesniveau is er een significant verschil, waarbij de vormgever significant meer feedback oplevert dan het verslag ($t(19) = -2,245$; $p = 0,037$, 95%CI[-0,22, -0,01]). In Tabel 3.7 staan de gemiddelde percentages weergegeven van de opmerkingen ingedeeld naar taak, proces, zelfregulatie en persoon.

Tabel 3.7 Percentuele verdeling van opmerkingen op de vormgever (n=9) en de verslagen (n=12) over verschillende feedback niveaus

Niveau	Verslagen			Vormgever		
	%	gem	sd	%	gem	sd
Taak	43,5	44	15	36	34	11
Proces	33	34	12	45	45	11
Zelfregulatie	23,5	22	11	18	17	9
Persoon	0	0	0	1	0	0

Feedback, feed up en feed forward

Wanneer de opmerkingen op validiteit geanalyseerd worden, dan blijken de meeste opmerkingen op basis van de formulering als feedback gecategoriseerd te kunnen worden (90%) van de opmerkingen. Veel minder opmerkingen kwalificeren als feed up (7%) of als feed forward (3%). De resultaten zijn in Tabel 3.8 weergegeven. Er lijkt echter niet direct sprake van misclassificatie: opmerkingen die in de rubriek feed up geplaatst worden hebben wel degelijk betrekking op doelen, criteria en verwacht niveau. In de formulering wordt daar weliswaar niet expliciet naar verwezen, maar in de begeleidende tekst bij de rubriek wel.

Tabel 3.8 Verdeling van opmerkingen op de vormgever over verschillende feedback categorieën (n=173). Hierbij is enkel gekeken naar de formulering; opmerkingen zijn daarnaast voorgesorteerd aan de hand van kopjes op de vormgever.

Categorie	%	gem	sd
Feed back	90	17,2	7,5
Feed up	7	1,4	1,6
Feed forward	3	0,6	0,7

3.6 Conclusie en discussie

In hoeverre leidt het gebruik van een vormgever voor feedback binnen de lerarenopleiding Biologie tot schriftelijke feedback volgens kwaliteitscriteria beschreven in de literatuur?

Op basis van de gegevens die beschikbaar zijn op het moment van schrijven kan er geconcludeerd worden dat het gebruik van de vormgever (in de bestudeerde context) leidt tot meer aandacht voor feed up en feed forward. Daarnaast lijkt er meer feedback op procesniveau gegeven te worden. Gekeken naar het feedbackmodel van Hattie en Timperley (2007) resulteert het gebruik van de vormgever dus in feedback die beter voldoet aan de kwaliteitscriteria. Daarmee ondersteunt deze casestudy de resultaten van Newton et al. (2012) en vult deze aan door de expliciete koppeling aan het model van Hattie en Timperley (2007). Andere kwaliteitscriteria zoals bijvoorbeeld beschreven door Nicol en Macfarlane-Dick (2006) zijn onderzocht aan de hand van een vragenlijst welke door studenten is ingevuld. Deze resultaten moeten nog geanalyseerd worden en zijn hier niet weergegeven.

Hoewel het vergelijken van de ene casestudy met een andere met voorzichtigheid betracht dient te worden lijkt het gebruik van de vormgever tot een verbetering te leiden ten opzichte van de situatie zoals beschreven in het voorgaande hoofdstuk. Ondanks variabelen, zoals andere docenten die betrokken zijn bij het onderzoek en verschillen tussen studentcohorten, sluiten de resultaten

van de twee casestudies nauw op elkaar aan. Aangezien de bevindingen in de voorgaande casestudy aansluiten bij resultaten uit de literatuur (zie hoofdstuk 2) lijken de gevonden resultaten betrouwbaar en bieden ze inzichten die mogelijk breder toepasbaar zijn.

Gebruik van de vormgever door de docenten

Uit de bestudeerde materialen en het criterium gericht interview (resultaten niet weergegeven) komt naar voren dat docenten de vormgever voor feedback als aanvulling gebruiken op de 'traditionele' manier van werken. De vormgever wordt dus gebruikt naast de ingevoegde opmerkingen in de tekst. Het levert daardoor nog niet de gewenste tijdswinst op.

Analyse van feedback op de vormgever

De vormgever is ingedeeld in vier categorieën: algemene indruk, feedback, feed up en feed forward. Gekeken naar de verdeling van opmerkingen over de laatste drie categorieën, blijkt er veel meer dan bij de opmerkingen bij de verslagen zelf, aandacht voor feed up en feed forward. Waar opmerkingen bij de verslagen vrijwel zonder uitzondering ingedeeld kunnen worden als feedback (97,5%), omvat de categorie feedback op de vormgever slechts 42% van de opmerkingen. 25% van de opmerkingen worden in de rubriek feed up geplaatst, 33% van de opmerkingen in de rubriek feed forward (zie Tabel 3.4).

Opvallend genoeg komt het bovenstaande resultaat niet terug wanneer er enkel tekstueel naar de opmerkingen op

de vormgever gekeken wordt (zie Tabel 3.7). Zo kan er in de rubriek feed up bijvoorbeeld een opmerking als *"In de methode van onderzoek mis ik het onderdeel waarmee je wilt kijken naar resultaten en kunt vergelijken of er een verbetering is opgetreden."* worden aangetroffen. De opmerking zegt op zichzelf niets over doelen, criteria of verwacht niveau, maar doordat de opmerking geplaatst is in een rubriek met als titel "Feed Up / essentiële aanpassingen" krijgt de boodschap een andere lading. Gekeken naar de inhoud lijkt er daardoor niet direct sprake van misclassificatie door de docenten. Mogelijk dat de onderverdeling van de vormgever in rubrieken er toe leidt dat er minder aandacht is voor het onder woorden brengen van feedback die direct te koppelen is aan doelen/criteria (feed up) of aan toekomstige doelen (feed forward). Een alternatieve verklaring is dat het formuleren van opmerkingen die expliciet herkenbaar zijn als feed up of feed forward erg lastig is zelfs na kennisname van het theoretisch kader. Dit laatste is ook beschreven in een publicatie door Walker (2007), waarbij de feedbackpraktijk onderzocht is, nadat docenten een handleiding voor het formuleren van effectieve feedback hadden ontvangen. Daarnaast is er beschreven dat ontwikkelingsgerichte feedback vaak impliciet geformuleerd is: duidelijk voor de betrokken docent, maar onduidelijk voor de betrokken studenten (Orsmond & Merry, 2011). Mogelijk dat het impliciete karakter van de feedback ook in de beschreven context geldt. In dat geval ondervangt de opbouw van de vormgever in verschillende rubrieken dit probleem doordat de titels op zichzelf al expliciet geformuleerd zijn.

Newton *et al.* (2012) beschrijven dat het gebruik van een vormgever ook leidt tot meer kwantiteit van feedback. Wanneer er gekeken wordt naar de hoeveelheid feedback die in de hier beschreven context wordt gegeven dan blijkt er inderdaad meer feedback gegeven te worden bij het gebruik van de vormgever. Voor zover er iets over gezegd kan worden bij de lage aantallen documenten die geanalyseerd zijn, werden er bij documenten die met de vormgever van feedback waren voorzien evenveel opmerkingen geplaatst als bij andere verslagen. Bovenop deze feedback kwam dan de ingevulde vormgever. Ook ten opzichte van een voorgaand onderzoek in dezelfde context werd er weinig verandering van het aantal opmerkingen zichtbaar. In dat voorgaand onderzoek zijn acht verslagen geanalyseerd met totaal 299 opmerkingen (*gemiddeld* = 37; *sd* = 19). Gekeken naar het aantal opmerkingen bij de negen verslagen die ook met de vormgever feedback hebben ontvangen (*n*=413; *gemiddeld* = 46; *sd* = 11) is er eerder sprake van een toename dan van een afname. Daar komen dan de opmerkingen op de vormgever (*n*=173, *gemiddeld*=19, *sd*=7) nog bij. Er is iets meer aandacht voor diepte 3 opmerkingen wanneer de tekst op de vormgever vergeleken wordt met de opmerkingen bij de verslagen (17%, respectievelijk 12,5% van de opmerkingen). Mogelijk dat dit gegeven te maken heeft met het feit dat de docenten van tevoren geïnformeerd zijn over het theoretisch kader van Hattie en Timperley (Hattie & Timperley, 2007).

Er wordt met name gebruik gemaakt van opmerkingen met diepte 1

Gekeken naar de diepte van de feedback kan vastgesteld worden dat het gebruik van de vormgever weinig invloed heeft. Opmerkingen met diepte 1 vormen veruit de grootste groep: 69% van de opmerkingen. Dit is vergelijkbaar met een eerder onderzoek in dezelfde context (67% van de opmerkingen). Opmerkingen waarin ook een uitleg of toelichting wordt gegeven vormen slechts een klein percentage van de opmerkingen (12%). Dit resultaat sluit aan bij eerder onderzoek (Glover & Brown, 2006; Walker, 2009). Wanneer de opmerkingen op de vormgever geanalyseerd worden op diepte, dan lijkt de verdeling niet zo sterk af te wijken van de opmerkingen bij de tekst. De grootste groep wordt gevormd door opmerkingen van diepte 1. Aan te brengen correcties worden nauwelijks genoemd (2%). Er lijkt ten opzichte van de opmerkingen bij de tekst misschien wat meer aandacht te zijn voor uitleg (12% van de opmerkingen in de tekst tegenover 17% van de opmerkingen op de vormgever).

Er wordt met name feedback gegeven op taakniveau

De grootste hoeveelheid feedback wordt gegeven op taakniveau. Dit gegeven zegt waarschijnlijk iets over het impliciet handelen van docenten, waarbij automatisch de taak centraal geplaatst wordt, maar het wordt daarnaast ongetwijfeld beïnvloed doordat het hier werk van studenten in de afstudeerfase betreft. In deze fase van de studie staat afstuderen immers centraal. Bij de formulering van feedback op de vormgever blijkt er meer aandacht te zijn voor proces, dus opmerkingen met betrekking tot de gekozen strategie van de student. Er lijkt wat minder

aandacht voor feedback op de niveaus van taak en zelfregulatie (niet significant).

Er wordt nauwelijks feed up of feed forward gegeven

In totaal kunnen 14 van de 554 opmerkingen bij de verslagen gerubriceerd worden als feed up of feed forward. Van deze 14 opmerkingen zijn er 11 geplaatst op verslagen van studenten die feedback hebben ontvangen *zonder* de vormgever. Van deze 11 opmerking werden er 10 gecategoriseerd als feed up en 1 als feed forward. Het lage aantal opmerkingen dat als feed up of feed forward aangemerkt kan worden bij de verslagen die ook middels de vormgever van feedback zijn voorzien is mogelijk te verklaren doordat docenten bij het formuleren van opmerkingen bij de tekst reeds rekening houden met de vormgever. Dit is niet verder onderzocht in deze casestudy. De lage percentages voor feed up en feed forward is ook bij een eerder onderzoek in dezelfde context vastgesteld (zie voorgaand hoofdstuk).

In de hier beschreven casestudy is de feedbackpraktijk bij een afstudeercursus in beeld gebracht door opmerkingen te categoriseren. Hoewel hier een beeld van de heersende praktijk uit naar voren komt, is het niet geheel duidelijk wat de getoonde getallen nu precies zeggen. Het is immers nog maar de vraag welke verhoudingen in type feedback het meest optimaal werkt en of dat voor verschillende studenten ongeveer gelijk is (zie ook Mayer & Alexander, 2011). Wanneer het leren van de student centraal geplaatst wordt, dan wordt het antwoord

op voorgaande vraag waarschijnlijk alleen maar meer complex. Immers, wanneer kan dat effect dan vastgesteld worden? En als er op een bepaald tijdstip geen effect gemeten kan worden, is het effect er dan niet? Of is er op het verkeerde moment gemeten? Deze problematiek wordt ook aangekaart door Price (2010). Daarnaast lijken er ook nog verschillen te bestaan in de optimale vorm van feedback afhankelijk van wat er geleerd dient te worden. Zo lijkt corrigerende feedback een effectieve methode te zijn wanneer het gaat om het aanleren van taal (Bitchener, 2008), terwijl corrigerende feedback geen informatie geeft over de 'gap' tussen de geleverde en gewenste prestatie (Glover & Brown, 2006). Mogelijk dat dit gegeven te maken heeft met de specificiteit van taak feedback (correcties vallen onder taak feedback). Taak feedback heeft namelijk als nadeel dat deze dikwijls dermate specifiek is dat deze niet bruikbaar is voor een ander curriculumonderdeel (Hattie & Timperley, 2007). In het geval van taalverwerving speelt dit mogelijk minder, aangezien juist taalgebruik in het overgrote deel van het curriculum bruikbaar is.

De bruikbaarheid van feedback vergroten kan dus een succeselement zijn, wanneer er gekeken wordt naar de effecten van feedback. Dit zou bijvoorbeeld onderdeel uit kunnen maken van een leerlijn, waarbij dezelfde vaardigheden verschillende keren terug komen in het curriculum. Op die wijze kan er mogelijk een groter effect verkregen worden van de feedback, terwijl er bovendien een kans ligt om feedback meer centraal te plaatsen in het curriculum en om studenten meer eigenaar te maken

van de ontvangen feedback (Price, et al., 2010). Onder meer vanuit deze gedachte is er bij de lerarenopleiding Biologie een onderzoeksleerlijn ontwikkeld met een belangrijke rol voor ontwikkelingsgerichte feedback en een terugkerende rubric voor beoordeling, geïnspireerd door onderzoek van Tractenberg (2009). Om het effect van feedback binnen deze leerlijn te bestuderen zal er een longitudinaal onderzoek uitgezet worden gedurende meerdere schooljaren. Daarnaast is het interessant om te bezien of er stappen gezet kunnen worden die leiden tot meer efficiënte feedback. Een van de oorzaken achter het geven van minder effectieve feedback is immers gebrek aan tijd (Hernández, 2012; Orrell, 2006; Weaver, 2006). Er zijn aanwijzingen dat het gebruik van een digitale omgeving kan helpen bij een tijdsefficiëntie (Gratama van Andel, 2014), dit facet kan hopelijk ook een plaats krijgen binnen het vervolgonderzoek.

To reduce the discrepancy between current and desired performance, the discrepancy can be reduced by:

Students

- Increased effort and employment of more effective strategies OR
- Abandoning, blurring, or lowering the goals

Teachers

- Providing appropriate challenging and specific goals
- Assisting students to reach them through effective learning strategies and feedback

Effective feedback answers three questions

Where am I going? (the goals)
How am I going?
Where to next?

Feed Up
Feed Back
Feed Forward

Each feedback question works at four levels:

Task level

How well tasks are understood/performed

Process level

The mental process needed to understand/perform tasks

Self-regulation level

Self-monitoring, directing, and regulating of actions

Self level

Personal evaluations and affect (usually positive) about the learner

Hoofdstuk 4

Effecten en resultaten van het inzetten van rubrics

Carel Burghout

4.1 Samenvatting

In dit hoofdstuk wordt beschreven hoe rubrics werden geïntroduceerd om te dienen als een instrument voor toetsing maar ook voor het geven en verwerken van feedback (self assessment en peer assessment) in twee Mastercursussen Engels bij Fontys Lerarenopleiding, Tilburg. Research betreft twee case studies met het doel te onderzoeken welke effecten en resultaten voortkomen uit het betrekken van studenten bij het maken van een rubric om hun werk te beoordelen.

In de eerste case study werd een rubric top-down geïntroduceerd in een cursus academic writing. In de tweede case study werd de rubric geschreven door de studenten, samen met hun docent. De uitslag van een onderzoek naar de bevindingen van studenten leidt tot de conclusie dat het betrekken van studenten door een discussie over hoe rubrics werken en hoe je met rubrics om kunt gaan twee positieve gevolgen kan hebben. Ten eerste werden studenten zich bewuster van de samenhang tussen hun leren en het beoordelen (constructive alignment). Daarnaast werden de betrokken studenten kritischer op hun eigen werk en voortgang, en daardoor

gemotiveerd. Voor de docent leidde het ook tot nieuwe inzichten wat betreft de complexiteit van de gestelde taak.

4.2 Inleiding

Het gebruik van rubrics is de afgelopen tien jaren sterk toegenomen in vele takken van onderwijs. Voorbeelden van rubrics voor allerlei vakgebieden zijn op het Internet te vinden. Studies naar het invoeren van rubrics bij toetsing van talenonderwijs werden aanvankelijk gestimuleerd door het wijdverspreide gebruik van rubrics om schrijfvaardigheid te helpen beoordelen en de laatste jaren door het steeds steviger erkende belang van het concept assessment for learning (Brennan, 2006; Sadler, 2005).

Arter en Chappuis (2006) beschrijven rubrics als een instrument dat de kenmerken definieert van werk van hoge kwaliteit. Deze kenmerken worden over het algemeen gevat in een tabel waarin de kwaliteitscriteria op verschillende niveaus zichtbaar wordt gemaakt. In analytische rubrics wordt elk apart criterium in de tabel gewogen in de uiteindelijke analyse van de kwaliteit van het werk.

In zogeheten ‘holistische’ rubrics zijn de criteria gegroepeerd in verschillende alinea’s die samen een totaalbeeld geven van hoe een bepaald niveau is bereikt of een bepaald product is ontstaan. Het oordeel wordt dan gebaseerd op een totaalinschatting van in hoeverre aan de criteria is voldaan. Minpuntjes in een deel van het proces of de taak kunnen worden getolereerd als de kwaliteit in zijn algemeenheid goed blijft.

Of ze nu holistisch of analytisch zijn, rubrics zouden een schaal voor het meten moeten zijn, en niet een lijst van inhouden of te nemen stappen. Zulke lijsten, aangeduid als ‘checklists’ komen de laatste tijd in zwang om gebruikt te worden naast de rubric. Burke beschrijft zulke checklists als ‘een routeplanner voor instructie’ die elke stap van een gecompliceerd proces of een complex geheel in te begrijpen eenheden of ondersteunende vaardigheden opdeelt. Daarnaast geven deze checklists de door de student of leerling te nemen achtereenvolgende stappen weer in termen van de te bereiken kwaliteitsstandaarden (Burke, 2011). Voor Burke zijn rubrics in vergelijking met deze checklists een ‘gids voor het scoren’ ontworpen om constructieve feedback te geven en de studenten te helpen helderder na te denken over de karakteristieke eigenschappen van werk van hoge kwaliteit. Tomlinson (2003) beschrijft rubrics als een kader ‘ontwikkeld om aan te tonen hoe belangrijke elementen van een taak er uit (zouden) zien in een continue progressie van “niet zo heel erg sterk ontwikkeld” tot aan “uitzonderlijk goed” (Tomlinson, 2003).

Rubrics blijken bijzonder goed bruikbaar te zijn als het gaat om het kijken naar complexe taken of het beoordelen van complexe vaardigheden, aangezien ze de criteria op een logische wijze rangschikken in een progressie van schalen van prestaties van onvoldoende naar uitstekend. Waar een checklist een goed hulpmiddel is om een taak aan te pakken door een blik op de delen geeft een rubric het geheel in zijn complexiteit weer en helpt zo studenten om vast te stellen hoe ze het op een bepaald moment doen en welke vervolgstappen ze dienen te nemen om tot een hogere kwaliteit te komen (Burke, 2008).

Het is belangrijk om bij het maken van een rubric eerst de specifieke criteria voor die taak specifiek te maken, dan daaruit observeerbare indicatoren af te leiden en dit alles in de tabel van de kwalitatieve schaal te vatten. Mertler (2001) stelt de volgende stappen voor bij het ontwerpen van rubrics voor het gebruik in een klas:

Eerst bekijk je de onderwijsdoelen nog eens goed, om daaruit observeerbare attributen af te leiden. Dan brainstorm je over karakteristieken die elk attribuut beschrijven. Vervolgens begin je met een verhalende beschrijving van het goede werk eerst, en dan wat je verstaan onder zwak of onvoldoende werk. Pas daarna vul je de andere vakken in het continuüm. Verzamel voorbeelden van werk die exemplarisch zijn voor elk niveau, deze worden je toekomstige benchmarks. Tenslotte moeten rubrics altijd geëvalueerd worden en zo nodig bijgesteld.

Aangezien de niveaus van een rubric refereren naar criteria en niet naar normen hoeft een score in het midden niet een gemiddelde of middelmagte score te zijn. Veel recente literatuur gaat niet in op scores en gewicht voor schalen maar wil zich liever richten op het inzetten van de criteria van rubrics als richtlijnen voor het geven van feedback (Andrade, 2000). In zulke gevallen worden rubrics gebruikt voor een formatieve beoordeling en zijn scores (nog) niet aan de orde. Wanneer rubrics worden gebruikt om het formatieve toetsen 'in lijn' te brengen met summatieve oordelen (zoals het geval is met de twee rubrics die in dit hoofdstuk worden besproken) dan kunnen scores wel relevant worden maar het vastleggen van vaste scores in de diverse niveaus brengt gevaren met zich mee; werken naar de rubric kan net zo vernauwend werken als alleen studeren gericht op de toets (Mabry, 1999).

Het gebruik van rubrics om taken te beoordelen heeft een lange geschiedenis en is de afgelopen dertig jaar wijd verspreid. Toch is het inzetten van rubrics ook stevig bekritiseerd. Bij talenonderwijs heeft het inzetten van rubrics om (creatief) schrijven te beoordelen op felle weerstand gestuit. In een vaak aangehaald artikel uit 1999 waarschuwde toets deskundige Linda Mabry dat rubrics de macht kunnen hebben om beoordeling juist te ondermijnen. Meer objectiviteit in beoordelen is het doel, maar grote standaardisering (en daardoor beperking) van een vaardigheid dat in zijn wezen individueel, expressief en creatief is kan het gevolg zijn. Daarnaast standaardiseren de rubrics ook het schrijfonderwijs, hetgeen het begrip van wat schrijven precies inhoudt en hoe je het kunt leren ondermijnt (Mabry, 2005).

De boodschap is duidelijk: wie rubrics introduceert om werk dat een zekere mate van creativiteit in zich moet hebben (bijvoorbeeld essays of het ontwikkelen van lesmaterialen) moet oppassen dat een ongewenste terugslag (backwash) kan ontstaan waardoor studenten de rubric hanteren om enkel te maken wat ze denken dat de docent van ze verwacht, en niets meer dan dat. Woorden als 'creativiteit' en 'individueel' toevoegen aan de rubric helpt niet tenzij deze termen in hun context worden besproken en verduidelijkt bij het presenteren van de rubric aan de studenten.

Er zijn echter ook andere stemmen in de discussie over het inzetten van rubrics. Andrade en Du (2005) bevelen rubrics sterk aan in de context van assessment for learning. Studenten zouden volgens hen uitgenodigd moeten worden om op de rubric te reflecteren, en te helpen met het maken van rubrics voor hun cursussen. Hun bevindingen zijn weer het uitgangspunt voor verdere studie van onder andere Reynolds-Keefer (2010). Reynolds-Keefer doet verslag over twee verschillende vormen van het inzetten van rubrics, het faciliteren van de metacognitieve processen bij studenten en het integreren van de cyclus van assessment in het onderwijzen en leren. Uit haar bevindingen blijkt dat rubrics op zich geen tovermiddel zijn als de rubric niet volledig begrepen wordt en er alleen echt sprake is van meerwaarde door het hanteren van de rubric bij gerichte feedback. Het betrekken van studenten bij het hanteren van de rubric is belangrijk, en helpt om onzekerheid over wat er precies gevraagd wordt en welke kwaliteit verwacht

wordt. Betrokkenheid bij de rubric stimuleert studenten om verantwoordelijkheid te nemen voor hun leren. Dit stimuleert op zich weer motivatie en helpt om het leren te verbeteren (Hattie & Timperley, 2007; Wulf, Raupach & Pfeiffer, 2005). Wanneer rubrics (mede) door studenten worden gemaakt wordt er op een hoog niveau gewerkt aan assessment for learning. Om een rubric met studenten te maken dienen docenten duidelijke leerdoelen te delen met hun studenten, dienen de studenten te worden voorzien van expliciete kennis van de inhoud. De docenten moeten een goed zicht hebben op de kennis van de studenten en assessment gebruiken om de instructie te verbeteren (Wilson, 2008).

4.3 De verlegenheidssituaties

De eerste verlegenheidssituatie diende zich aan in 2011. Docenten wilden op één lijn zitten met collega's wat betreft de uiteindelijke beoordeling van essays. Daarnaast wilden docenten de beoordeling van essays transparant maken. Tenslotte bestond er ook een behoefte aan een instrument dat zowel formatief als summatief kon worden ingezet. Docenten merkten dat de feedback die werd gegeven op de eerste opzetten van de essays en ook op het uiteindelijk geleverde werk werd genegeerd of niet goed werd verwerkt. Dat betekende dat de inbreng van de docenten in het uiteindelijke resultaat veel te groot werd, aangezien zij dan de neiging hadden om 'de boel recht te zetten'.

Studenten dienden een beter beeld te krijgen van wat er uiteindelijk van ze verlangd werd en aangespoord te worden om zelf correcties aan te brengen voordat ze het werk inleverden. Een rubric 'Academic Writing' leek daarvoor een geëigend instrument.

De tweede verlegenheidssituatie rees vlak daarna. In een andere cursus (20^e-eeuwse poëzie) vroeg de beoordeling van dossierwerk en didactisch werk voor poëzie om een groot aantal verschillende, in hun samenhang complexe vaardigheden. Daarbij rees de vraag, hoe de studenten te verleiden tot zelfsturing in de selectie van de gedichten waarmee zij die vaardigheden goed voor het voetlicht konden brengen. Welke invloed zou het hebben als de rubric voor het werk niet door de docent werd gemaakt? Was het mogelijk om een rubric door de studenten te laten ontwikkelen op basis van de criteria die genoemd werden in de Kennisbasis Engels voor de Masteropleidingen en in de studiegids? Het werken aan de standaarden voor zo'n rubric zou de studenten dan mogelijk een goed inzicht verschaffen. Jonsson and Svingby (2007) benadrukken dat rubrics studenten helpen om criteria te begrijpen, feedback versterken en mogelijkheden bieden voor peer- en self assessment.

4.4 Onderzoeksvraag

Daaruit kwam de volgende onderzoeksvraag voort: *Welke effecten heeft het invoeren van werken met rubrics op het assessment, zowel formatief als summatief, van essays en werkstukken op Master niveau Engels?*

Tijdens het onderzoek rees na het doorspreken van de doelen en de werkwijzen met de studenten de sub vraag: *Welke gevolgen heeft het betrekken van de studenten bij het maken van de rubrics voor inzicht in de leerdoelen en de te nemen stappen om ze te bereiken, en motivatie om een goede of excellente prestatie te leveren?*

4.5 Beoogde resultaten

De bedoeling was dat de studenten in beide gevallen meer zicht kregen op de afzonderlijke onderdelen die leidden tot de complexiteit van de opdracht.

Daarnaast moesten de duidelijk geformuleerde criteria en leerdoelen, uitgedrukt in indicatoren, de studenten in staat stellen om zelfstandig, met de rubric, verbeteringen aan te brengen.

Ook werd beoogd dat de rubric de student zou helpen om sterkten en zwakten te verwoorden zodat de docent de juiste (soort) feedback kon geven op het goede moment. Tenslotte kwam er nog een doel bij: dat het samen met studenten ontwikkelen van een rubric zou motiveren (inzicht geven in self-efficacy) en leiden tot een kritische houding ten opzichte van het te leveren werk.

4.6 Methode: twee pilot cases

Het onderzoek naar effecten en resultaten van het inzetten van rubrics bestrijkt een periode van 3 studie jaren. In de drie opeenvolgende jaren (2011-2014) zijn rubrics geïntroduceerd in cursussen voor de Masteropleiding Engels aan Fontys Lerarenopleiding Tilburg.

De studenten zijn op verschillende manieren bevraagd over de invloed die de rubrics hadden op de aanpak van hun essay- en dossier opdrachten en op de kwaliteit van hun werk. Na een top-down begin zijn de studenten steeds meer betrokken bij het formuleren van de standaarden voor de rubrics. Daarnaast zijn de rubrics steeds sterker ingebed als een instrument voor feedback zowel als voor beoordeling.

Het literatuuronderzoek leverde gaandeweg nieuwe inzichten op, net als het werken met de rubrics met de studenten. Dat had onder andere invloed op het moment waarop de rubrics geïntroduceerd werden en de vragen die gesteld werden, bijvoorbeeld of de bekendheid met rubrics leidde tot het frequent raadplegen van de rubric bij het werken aan een essay of een dossier. Studenten werden ook gevraagd aan te geven op welke manieren de rubric hen had geholpen om werk van een hogere kwaliteit te leveren omdat ze hun aanpak konden vergelijken met vergelijkbaar werk zonder rubric.

Verwachtingen

De bedoeling was dat de studenten in beide gevallen meer zicht kregen op de afzonderlijke onderdelen die leidden tot de complexiteit van de opdracht.

Daarnaast moesten de duidelijk geformuleerde criteria en leerdoelen, uitgedrukt in indicatoren, de studenten in staat stellen om zelfstandig, met de rubric, verbeteringen aan te brengen.

Ook werd beoogd dat de rubric de student zou helpen om sterkten en zwakten te verwoorden zodat de docent de juiste (soort) feedback kon geven op het goede moment.

Tenslotte kwam er nog een doel bij: dat het samen met studenten ontwikkelen van een rubric zou motiveren (onder andere door verkrijgen van inzicht in self-efficacy) en leiden tot een kritische houding ten opzichte van het te leveren werk.

Context

De onderzoeksvraag is geformuleerd in het studiejaar 2011-2012. Bij het denken over en werken aan een transparante opzet voor alle toetsing aan de Master opleiding Engels leidde literatuuronderzoek naar de kennis en vaardigheden die wij willen beoordelen in literatuur- en cultuur onderwijs naar de mogelijkheden die worden geboden door rubrics. Rubrics zijn geschikt om in te zetten bij het beoordelen van complexe vaardigheden, of producten die die complexiteit weerspiegelen.

Essays en (didactische) werkstukken en dossiers op master niveau zijn zulke producten. Ze vormen deel van het toets repertoire aan de Masteropleiding Engels aan Fontys

Lerarenopleiding Tilburg. Er waren echter in 2011, in tegenstelling tot de beoordeling van het praktijkonderzoek binnen dezelfde opleiding, geen kaders voor beoordeling van die producten.

Het onderzoek heeft zich uiteindelijk uitgestrekt over de studie jaren 2011-2012, 2012-2013 en 2013-2014, respectievelijk jaargang 1, jaargang 2 en jaargang 3 genoemd.

In de eerste pilot case, in het studiejaar 2011-2012 kregen de studenten (jaargang 1) een door de docenten ontwikkelde analytische rubric voor Academic Writing. Bij het samenstellen van die rubric werd uitgegaan van toepasbaarheid in alle gevallen waarin studenten een academisch essay moesten produceren, de rubric was dus niet specifiek voor die cursus alleen. Deze groep bestond uit 18 studenten

Studenten werd gevraagd de rubric goed door te nemen, te vragen waar eventuele onduidelijkheden zaten en vervolgens om de rubric toe te passen op hun werk in ontwikkeling voordat ze dat werk overlegden aan docenten. Het werd van meet af aan duidelijk gemaakt dat de rubric die ze op deze manier formatief inzetten (self-assessment van werk in ontwikkeling) ook gebruikt zou worden bij de beoordeling.

In de tweede pilot case kregen dezelfde studenten een half jaar later de criteria (standards) aangereikt na een college waarin rubrics en hun eigenschappen werden besproken.

Na een week dienden zij criteria te hebben uitgewerkt en indicatoren voor de rubric te hebben gemaakt. Met wat er werd aangeleverd werd vervolgens in overleg met de hele groep (jaargang 1) de eerste versie van de rubric voor de cursus poëzie in elkaar gezet.

In beide gevallen werd de studenten gevraagd om feedback over de effecten van het gebruiken van de rubrics. Dat was aanvankelijk in de vorm van een enquête formulier, later, om enquête 'moeheid' te voorkomen, in de vorm van interviews.

4.7 Resultaten

In de eerste case, Academic Writing, waren de effecten al heel snel waarneembaar. Uit de respons op een enquête halverwege de cursus (12 respondenten) bleek het volgende:

1. Studenten ijkten werk eerst aan de rubric, om eventueel te verbeteren, om dan pas in te leveren.
2. Er was minder benauwdheid omtrent het schrijven aangezien men beter wist wat er precies verwacht werd.
3. Er kwamen minder vragen over hoe het werk beoordeeld werd.
4. Er was veel meer aandacht voor de feedback dan voorheen en studenten gaven aan dat ze zich meer inspanden om het werk te structureren.
5. Studenten gaven in de enquêtes aan dat er een grote acceptatie van het instrument was en een wijder gebruik van rubrics in de eigen lespraktijk

(de studenten zijn allemaal al leraar). Kanttekening: het zijn kleine groepen (ongeveer 18 studenten en daarbinnen nog een 2/3 respons, er zijn geen brede conclusies mogelijk).

In de tweede case, Poëzie Dossier, viel het op dat de groep al op goed niveau kon meepraten over rubrics en zelf slaagde in het formuleren van de meeste indicatoren. Waar dat niet lukte werden ze door de docent aangevuld. De studenten gebruikten de rubric voor de samenstelling van hun dossier en het geven van feedback aan hun peers. De docent gebruikte de rubric voor feedback en het uiteindelijke oordeel.

De 2e jaargang in de cursus Poëzie (cursisten studiejaar 2012-2013) bracht een belangrijke verbetering aan in de structuur van de rubric en gaf feedback op waar het taalgebruik helderder en eenduidiger kon. Uit de enquête kwam een, ten opzichte van de groep in de eerste case, minder algemeen positief beeld van de invloed van de rubric op het geleverde werk. Maar daar stond tegenover dat zij aangaven dat het meedenken over de standaarden die hun werk moest behalen zeer motiverend was.

De 3e jaargang (cursisten studiejaar 2013-2014) zag geen reden om de verbeterde rubric nog verder te veranderen of aan te passen. Dit was een inschattingsfout, de opdracht was eigenlijk niet meer bottom-up, een volgende jaargang zou wederom de opdracht moeten krijgen om zelf een rubric te bouwen om de effecten vergelijkbaar te doen zijn. Met de 3e jaargang werd vervolgens de

rubric toegepast in een peer assessment van het werk in aanbouw van een medestudent. De student kon zo de feedback meenemen in het vervolg van zijn werk en de andere studenten konden de inzichten die boven kwamen goed gebruiken bij het samenstellen van hun eigen dossier. Uit dit werk kwam een verdere standaardisatie van het werken met de rubric voort: studenten dienen voortaan halverwege de cursus hun werk tot dan toe te presenteren in een korte voordracht, waarbij ook de koppeling met het werken met de rubric in al zijn facetten aan bod komt.

4.8 Conclusies en discussie

Conclusies uit de enquêtes en gesprekken met studenten leveren het volgende beeld: er is waardering voor de inzet van het instrument rubric. Studenten stellen dat ze bewuster omgaan met feedback. Daarnaast geven ze aan dat het bespreken van de rubric en het werken ermee helpt om bewuster te worden (in de eigen lespraktijk en ook als student) van het belang van transparante, objectieve en betrouwbare beoordeling (double loop learning).

De docent ervoer dat de effecten van het inzetten van een rubric in de eerste case leidde tot hogere kwaliteit van het ingeleverde werk en een grotere transparantie van de feedback en het eindoordeel. Daarnaast heeft de docent zelf het belang geleerd van de ‘vertaalslag’ van ‘criteria c.q. doelstellingen samen opstellen met studenten’. Tenslotte heeft de docent een beter beeld gekregen van de complexiteit van de gestelde taken.

In zijn overzichtsartikel over Assessment for Learning stelt Dylan Wiliam dat assessment processen grote invloed hebben op de bereidheid maar ook capaciteit om te leren (Wiliam, 2011). In hun review van studies naar de effecten van rubrics stellen Anders Jonson en Gunilla Svingby vast dat het potentieel van rubrics om het leren en het onderwijzen te versterken voornamelijk ligt in het feit dat rubrics verwachtingen en criteria expliciet maken en zo feedback en self-assessment faciliteren (Jonson en Svingby, 2007). De uitkomsten van dit kleine onderzoek sluiten aan bij de conclusies van deze veel bredere studies en nodigen uit tot verdere inzet van rubrics bij complexe taken en verdere studie naar de effecten daarvan. Kay Burke beveelt het gebruik van checklists in combinatie met rubrics aan (Burke 2001), Panadero een combinatie van rubrics en scripts (Panadero, Alonso Tapia, & Huertas, 2012). Beide aanbevelingen leiden weer tot nieuwe vragen en de mogelijkheid om de effectiviteit van zulke combinaties te onderzoeken.

Terugblik op de kenniskring door Carel Burghout

Allereerst is het voor mij buitengewoon leerzaam geweest om te ervaren hoe goed en serieus deze studenten kunnen meepraten over standaarden en hoe belangrijk het is om aan het begin van een cursus discussie te hebben over standaarden. Het meewerken aan de rubric verschaftte ook mij inzicht in de complexiteit van de gestelde taak en droeg daarmee bij tot het gerichter kunnen geven van feedback. Het doen van onderzoek en er over schrijven heeft mij ook doen inzien hoe lastig het is om wat voor mezelf evident was voor andere mensen helder en samenhangend over te brengen. De positieve effecten van de inzet van rubrics (ver)leiden tot het uitwisselen van ervaringen met collega's aan de eigen opleiding en andere opleidingen. De literatuurstudie heeft mij van een focus op rubrics gebracht naar het veel breder lezen over Assessment for Learning. Er is belangstelling van collega's en van andere opleidingen om geschoold te worden in het ontwerpen van en gebruiken van rubrics. De meerwaarde van het werken in een kenniskring uit zich in de voor de hand liggende koppeling naar onderzoek naar de effecten van kwalitatief hoogwaardige feedback en naar de koppeling van het inzetten van rubrics samen met checklists en / of scripts.



Rubric worksheets Modern Novel M Ed FLOT 2011 Version 3				
Feature	Insufficient	Sufficient	Good	Excellent
Design, inviting Elicits reader response, fun to do	No or few appeals made to senses by design, or chaotic	Design is attractive but not always relevant	Attractive design draws in attention with relevant images /layout	Excellent design, draws in attention, images are integrated, higher purpose than just visual effect or decoration
Both general enough and specific enough	Could relate to various novels, or covers just one chapter / incident	Covers a relevant section or theme or series of incidents in the novel	Items clearly relate to a specific theme or topic and cover this novel adequately	Chosen approach and topic(s) all relate to theme(s) of the novel and interrelate
Focus: related set of themes, incidents, symbols, etc	A list of questions with insufficient or no focus	Most questions focus on a relevant section or theme or series of incidents in the novel	Relevant focus, sustained throughout. Clearly related items	Relevant and tight focus with varied approaches all leading back to better understanding
Scaffolding: help and stretching to ZPD	No apparent scaffolding or scaffolding set too low or too high	The learner is helped and challenged occasionally	The steps are consistent in both helping and challenging	Well thought-out series of steps take learner to higher understanding
What it potentially shows the teacher	Shows memorising, no new or deeper learning made visible	Shows understanding of the novel at the expected level	Brings out the learner's experience and enjoyment of the novel as well as good understanding	Shows understanding, enjoyment and experience of the book leading to new / deeper learning
Easy to mark, easy to rate?	Many (sub)questions, no clear weight given to items, too open	Same questions show clearly what is expected but some still too open, no weighting of items	It is quite clear what is expected but no weighting of items	WS makes clear what is expected and what will count

Hoofdstuk 5

Praktijkonderzoek naar inzet van rubrics, checklists en scripts bij vakdidactiek

Monique Maas

5.1 Samenvatting

Dit praktijkonderzoek richt zich op het verbeteren van het verwerven van toepassingsgerichte vakdidactische kennis bij deeltijdstudenten in de lerarenopleiding Engels. De moeilijkheid voor studenten zit met name in de transfer van theoretische kennis naar toepassingsgerichte kennis. Dat wil zeggen in het correct en adequaat vertalen van allerlei vakdidactische mogelijkheden naar de klassensituatie op de (stage-)school van de cursist. Het maken van gefundeerde vakdidactische keuzes, waarbij leerdoelen gerealiseerd worden in een effectieve activerende didactiek en waarin werkvormen, leermaterialen en toetsing goed op elkaar aansluiten, vergt veel van lerenden.

Vanuit de literatuur zijn verschillende instrumenten beschreven die studenten bij doelmatig(er) leren kunnen ondersteunen. Dat zijn met name de combinatie van rubric en scripts. Deze instrumenten zijn toegepast

binnen verschillende cursussen vakdidactiek waarbij het aanbod werd aangevuld met gerichte klas-activiteiten en werkvormen. Het praktijkonderzoek liet zien dat inzet van met name rubrics en scripts studenten inderdaad in staat stelt om doelmatig toepassingsgerichte kennis op te doen en hun eindproducten (o.a. lesmaterialen) te laten onderbouwen vanuit de theorie. Voorwaarde is wel dat in de lessen nadrukkelijk aandacht aan omgang met de rubric en scripts wordt besteed en gedurende de volgende bijeenkomsten de ondernomen activiteiten telkens terugverwijzen naar deze instrumenten, bijvoorbeeld in peer-assessment werkvormen.

5.2 Inleiding

Dit praktijkonderzoek richt zich op het verbeteren van het onderwijs vakdidactiek aan deeltijdstudenten. Uit verschillende publicaties blijkt dat het aanleren van complexe vaardigheden goed ondersteund kan worden

door gebruik te maken van rubrics (Arter & Chappuis, 2006; Jonsson et al, 2007; Panadero, 2012) aangevuld met checklists (Burke, 2011) en scripts (Panadero et al., 2012). De transfer van vakdidactische kennis naar toepassingsgerichte kennis bij studenten die vaak gelijktijdig vakdidactiek cursussen volgen en stage lopen kan aanzienlijk effectiever vormgegeven worden door inzet van die instrumenten en die werkvormen. Studenten leren gefundeerde vakdidactische keuzes te maken, om beter beslagen ten ijs te komen tijdens de stages en om grotere leerwinst te boeken bij het verwerven van toepassingsgerichte vakdidactische kennis, waar zij zonder die hulpmiddelen soms moeite hebben met het verwerven van vakdidactiek en moeite hebben om opgedane kennis vervolgens te vertalen naar hun beroepspraktijk. In het onderzoek worden drie instrumenten ingezet die het bewuster leren van de student mogelijk maken doordat deze instrumenten lesdoelen en streefniveaus duidelijk maken via de ingezette rubric. Een rubric is een schematisch overzicht van de beoordelingscriteria waar de leerdoelen van de cursus gekoppeld worden aan kwalitatieve niveaubeschrijvingen en voorbeelden van het geleverde werk (Jonsson et al, 2007; Burke, 2011). Via checklists zijn de complexe opdrachten in kleinere eenheden gevat. Als aanvulling op het gebruik van de rubric zijn checklists per opdracht gebruikt als hulpmiddel om complexe, abstracte en afzonderlijke taken in een behapbaar stappenplan op te knippen. De checklist verduidelijkt afzonderlijke eisen en kadert de opdracht nauwkeuriger in. Het is tevens een uitwerking van de

opdrachten die opgenomen zijn in de rubric. Checklists zijn ingezet als controlelijsten voor studenten.

Waar in de rubric staat: 'Appropriate books for *onderbouw* / Two work sheets for each book. / Teacher guide linking theory to work sheets and objectives', bevat de checklist veel uitgebreidere informatie: "Focus on finding two appropriate books for a lower form (*onderbouw*) extensive reading programme, explaining genre and level in relation to your intended target group. Develop a work sheet for each book in which you combine clear target with stimulating assignments in which students explore two or three aspects of a reader. Never losing the ultimate aim of language development and positive reading experiences that stimulate further reading out of sight."

Ten slotte zijn scripts ingezet die studenten stapsgewijs helpen bij het vervullen van de complexe opdrachten. Scripts zijn voorgestructureerde, stapsgewijze vragen die tezamen een stappenplan opleveren vanuit een expertmodel, over hoe een complexe taak uit te voeren in een complexe beroepssituatie (Panadero et al., 2012).

5.3 Onderzoeksvraag

In hoeverre kunnen rubrics, scripts en checklists een bijdrage leveren aan het ontwerp van een doelmatig leertraject voor vakdidactiek bij Engels voor deeltijders, waarbij de nadruk ligt op een effectieve transfer van theoretische vakdidactische kennis richting toepassingsgerichte kennis?

5.4 Methode

Context

Sinds ongeveer een jaar wordt praktijkonderzoek uitgevoerd bij deeltijd cursussen vakdidactiek voor eerstejaars bij Bachelor studenten Engels (FLOT). Zowel volwassenen in de deeltijd opleiding en adolescenten in de voltijd variant van de opleiding tot docent Engels worstelen met het correct toepassen van het geleerde in een complexe beroepspraktijk. Het verwerven van de theorie van vakdidactiek is veelal geen probleem. De moeilijkheid van de toepassing zit met name in de transfer van theoretische kennis naar toepassingsgerichte kennis. In deze context betekent dat in het correct en adequaat vertalen van allerlei vakdidactische mogelijkheden naar de klassensituatie op de (stage-)school van de cursist. Diversiteit, niveauverschillen, groepsdynamiek, leerstijlen en verschillen in voorkennis vragen om afgewogen vakdidactische keuzes. Het maken van deze gefundeerde vakdidactische keuzes, waarbij leerdoelen gerealiseerd worden in een effectieve activerende didactiek en waarin werkvormen, leermaterialen en toetsing goed op elkaar aansluiten, vergt veel van lerenden.

Ervaren docenten maken op basis van hun expertise razendsnel keuzes welke werkvormen (op basis van de leerdoelen) bij welke leerstof horen. Zij weten hoeveel tijd hiervoor in te ruimen en hoe te evalueren, etc. Het vergaren van die expertise is de grote uitdaging voor een student vakdidactiek. Uit onderzoek blijkt dat het aanleren van complexe vaardigheden goed ondersteund

kan worden door gebruik te maken van rubrics (Arter & Chappuis, 2006; Panadero, 2012).

Instrumenten

Door middel van een self-efficacy vragenlijst is gemeten of de inzet van rubric, scripts en checklist het leren van studenten doelmatiger maakte in twee eerstejaarsgroepen van ca 20 studenten. De antwoorden op de vragenlijst zijn vergeleken met toetsresultaten, observaties tijdens de les, periode-evaluaties en interviews met studenten.

In cursus 1.4 in 2013 (vakdidactiek van grammatica en vocabulaire) is een self-efficacy vragenlijst ingezet afkomstig van Van Meeuwen (Van Meeuwen, Brand-Gruwel, Kirschner, De Bock en Van Merriënboer (2012)) die meet of studenten vertrouwen hebben in het goed vervullen van hun studietaak.

De vragen zijn licht aangepast, het woord 'taak' is vervangen door 'cursus', bijvoorbeeld:

	zeer oneens	oneens	oneens/ eens	eens	zeer eens
1 Ik ben er van overtuigd dat ik de basisvaardigheden die deze taak traint, zal leren.	☐	☐	☐	☐	☐

is geworden:

	zeer oneens	oneens	oneens/ eens	eens	zeer eens
1 Ik ben er van overtuigd dat ik de basisvaardigheden die deze cursus traint, zal leren.	☐	☐	☐	☐	☐

In het artikel van Panadero wordt onderzocht of studenten zelfstandig en effectief leren doordat zij beschikken over rubrics, checklists en doordat zij geholpen worden door scripts. Om te meten of de deelnemers van de deeltijd vakdidactiek cursus 1.4 in staat waren om met vertrouwen hun eigen leerproces vorm te geven (self-efficacy), is een vragenlijst ingevuld, die drie variabelen onderzoekt: *learning*, *task performance* en *task value*.

De vragenlijst is afgenomen op de laatste bijeenkomst onder alle deelnemers in beide deeltijdgroepen, in totaal 25 deelnemers, van vakdidactiek 1.4. De antwoorden op de vragenlijst gaven aan dat de deelnemers van vakdidactiek 1.4 geloven dat zij kunnen voldoen aan de gestelde opdrachten (Self-efficacy for task performance).

Procedure

Als eerste is een rubric ontworpen, omdat dit instrument helderheid en structuur aanbrengt en de eisen expliciteert. Daarna zijn checklists (Burke, 2011) geformuleerd. Vervolgens zijn scripts (Panadero, 2012) ingezet. Tijdens de eerste cursusbijeenkomst worden de eisen verwoord in de rubric en checklists uitgebreid doorgenomen, alsmede de criteria waaraan het portfolio moet voldoen.

Tijdens de lessen zijn er wekelijks formatieve feedback momenten waarbij (delen van) het te ontwerpen lesmateriaal worden besproken (*double-loop* principe).

Tijdens de cursus is middels groepswork, voorafgaand aan bepaalde uit te voeren opdrachten (bijv. het maken van lesmaterialen en geven van mini-lessen) door studenten samengewerkt om stappenplannen op te stellen bij opdrachten, die vervolgens plenair besproken en vervolmaakt werden. Deze scripts zijn tenslotte op de digitale leeromgeving geplaatst ter ondersteuning bij het maken van de opdrachten thuis.

Een voorbeeld van een script

Script

Intensive reading lesson

9 December 2013

Leerdoel + werkvorm / activiteit afstemmen

- waar vind je de leerdoelen? Zie Staatsen
- maak een keuze! Niet alle doelen in één les proppen

Doelgroepaansprekend material kiezen

Beginniveau + beginsituatie inschatten en meenemen in planning

- Wat is er al behandeld?

Rekening houden met groepsdynamiek

- veilige situatie in de klas / ordeproblematiek?
- verschillende rollen en perspectieven in acht nemen bij planning (vanuit docent en vanuit leerling)

Hoe evalueer je? Wat meet je?

-link hier een afsluitende werkvorm aan (bijv. een kleine presentatie)

Neuner typologie in les oefeningen, van A naar D stapsgewijs**Actueel, authentiek materiaal (wanneer mogelijk)**

Een bekend nadeel van werken met rubrics en scripts is het gevaar van *teaching to the test*. Door de eindopdrachten te construeren als complexe, open opdrachten met meerdere correcte oplossingen is dit mogelijke nadeel ondervangen. Zo moeten studenten bijvoorbeeld lesmateriaal voor intensief lezen voor de onderbouw ontwerpen. Daarnaast moet een *teacher guide* bijgevoegd worden, waarin de lesopzet verantwoord wordt vanuit de theorie en ingegaan wordt op de keuzes die zij voor hun (stage) doelgroep hebben gemaakt.

5.5 Resultaten

De eindevaluaties, anoniem afgenomen tijdens het laatste college, gaven aan dat studenten baat hadden bij de ingeslagen weg waarbij rubrics, checklists en scripts ingezet worden. Zij verwoordden dat zij inderdaad doelmatiger door de stof heen komen en beter de "linkage" bereiken tussen theorie en toepassingsgerichte kennis op vakdidactiekgebied. Hieronder is een impressie

van enkele vragen en verkregen studentantwoorden ter illustratie te lezen.

Heb je tijdens de cursus en bij het studeren gebruik gemaakt van de rubric? Zo nee, waarom niet? Zo ja, waarom wel?

"steeds -> steeds controleren of je de juiste lijn volgt."

"Ja concreet en duidelijk overzicht van wat er nu van me verwacht wordt."

"Ja, in periode 3 [4, sic] meer gebruikt omdat opzet nu duidelijk was."

"Wel gebruikt, niet zo overzichtelijk schema."

Welke concrete kritische leervraag of leervragen uit de rubric heeft je inzicht gegeven in jouw leerproces of heeft je geholpen bij het voldoen aan de eisen?

"Meet je wat je wilt meten?"

"Sluit materiaal en werkvorm aan bij de doelgroep."

"Is het materiaal geschikt voor het niveau?"

"De (kern)doelen + Neuner typologie zodat ik steeds leer verantwoord wat ik doe en een duidelijk geleide opbouw leer toe te passen in mijn opdrachten."

"Het feit dat je de inhoud (incl. doelen) van de module kunt uitleggen aan de ander -> goede controle. Checklists handig en kijken naar andere! Feedback geven en krijgen."

"Nvt"

Was de opzet van de cursus vakdidactiek 1.4, waarbij we in de colleges stapsgewijs de cursuseisen behandelden en oefenden om de vakdidactiek t.a.v. vocabulaire en grammatica onderwijs te leren doorzien en toepassen goed vormgegeven? Licht je antwoord toe. Ik vond het nuttig en leerzaam, maar ook pittig. Ik mis met name voorbeelden van een teacher guide."

"Ja minder verwarrend dan 1.2, meer begeleiding gehad, duidelijk en stapsgewijs."

"Ja eerst theorie, dan regels van de opdracht, dan toepassen, daarna bespreken in de klas en dan puntjes op de i."

"Ja, maar dat kwam mede door de cursus vakdidactiek 1.2, omdat daar de opzet uiteindelijk duidelijk werd wist ik beter wat er nu in 1.4 van me werd verwacht en waar naar toe gewerkt moest worden, m.a.w. eenmaal gewend -> prima!"

Helpt de checklist (de lijst met eisen die je aantreft in het programma voorafgaand aan de rubric) t.a.v. de vocabulaire en grammatica en trouble-shooting zaken die in het portfolio opgenomen moeten worden jou om de leerdoelen te bereiken?

"Ja, het maakt een stapsgewijze aanpak mogelijk."

"Ja, dat is mijn houvast/afvinklijst voor het maken van het portfolio."

"Nee, nog veel onduidelijk."

Zijn de scripts (de vragenlijsten die we tijdens de les plenair hebben opgesteld en die stapsgewijs de te maken keuzes verhelderen bij bijvoorbeeld het samenstellen van de deductieve grammatica instructie, of het samenstellen van de inductieve grammatica les) behulpzaam? Licht je antwoord toe.

"Yes, because all questions were answered before we made any mistakes through misunderstanding and ensured we were all on the same page from day one."

"Ja, dit is eigenlijk het verlengde van de rubric, maar dan klassikaal behandeld."

"Scripts?"

"Kan ik me niet herinneren."

Kun je in eigen woorden uitleggen waarom je wel/geen gebruik hebt gemaakt van de rubric en/of de checklist en/of de scripts?

"Ik gebruik de (vragen)lijsten & eisenlijst bij het maken van

de opdrachten om de juiste vragen te stellen. Wat nou wat is (rubric/script/checklist) vind ik onduidelijk en eigenlijk ook niet zo relevant. Wellicht is 1 document handiger?"

"Ja. Omdat je met het gebruik van deze hulpmiddelen, een zo goed mogelijke les kan voorbereiden."

"Ja, wekelijks"

De scores van de self-efficacy vragenlijsten wijzen op vertrouwen bij deelnemers dat zij in staat zullen zijn om bepaalde opdrachten te kunnen uitvoeren op het gevraagde niveau (Self-efficacy for learning), namelijk gemiddeld 3,6 op een schaal van 1 tot 5.

Bij de Self-efficacy for task value, waarvan de scores wijzen op een groot geloof in het nut en de relevantie van de opdrachten voor hun latere beroepspraktijk, is de gemiddelde score 3,8 op een schaal van 1 tot 5.

De betrouwbaarheid van de vragenlijst is gecontroleerd door het uitrekenen van de Cronbach's Alpha. Voor de self-efficacy for task performance bleek de interne consistentie 0,752, voor de self-efficacy for learning vragen bleek die 0,632, wat waarschijnlijk te wijten is aan een negatief gestelde vraag, wanneer die vraag 6 weggelaten wordt is de score 0,705.

De interne consistentie voor self-efficacy for task value vragen is 0,794.

5.6 Conclusies en discussie

Uit observaties tijdens de 8 bijeenkomsten en een analyse van de toetsresultaten alsmede uit de respons op de vragenlijsten en de evaluaties blijkt dat de gecombineerde inzet van rubric, scripts en checklists werkt en effectiever

leren van studenten mogelijk maakt. Studenten leveren goede portfolio's in met bruikbare lesmaterialen en zijn in staat om hun werk met theoretische inzichten te onderbouwen.

Studenten waarderen inzet van scripts, rubrics en checklists, maar moeten goed geholpen worden bij het doorzien van de instrumenten. Dit vergt inspanning en lestijd en aparte werkvormen. De rubric is richtinggevend voor studenten, maar ook voor docenten, want lestijd inruimen voor werken met de instrumenten is randvoorwaardelijk. De vormgeving van de rubric met name moet duidelijker in de toekomst.

Het gebruiken van rubrics in combinatie met scripts en checklist levert betere producten op, studenten blijken in staat om effectieve materialen voor leerlingen te ontwerpen en vanuit de theorie te kunnen onderbouwen.

Verbeterpunten in het uitvoeren van het praktijkonderzoek zitten in een inzet van beter doordachte meetinstrumenten waardoor het duidelijker wordt waar het positieve effect precies door veroorzaakt wordt. Met name het verder uitwerken van de veelbelovende werking van scripts en rubrics in combinatie met activerende didactiek en peer-assessment opdrachten tijdens de bijeenkomsten is een speerpunt in volgende praktijkonderzoeken. Het is niet vanzelfsprekend dat studenten de instrumenten meteen toepasbaar vinden. Tijdens de bijeenkomsten moeten de instrumenten en de activiteiten worden afgestemd. De nadruk moet in de eerste bijeenkomst liggen op het elkaar

uitleggen hoe de middelen gebruikt kunnen worden. Gaandeweg zal in de bijeenkomsten telkens terug moeten worden verwezen naar de rubric en de checklist om bijvoorbeeld in peer-assessments elkaars werk te bespreken en bij te stellen. Geconcludeerd kan worden dat studenten de inzet van scripts, rubrics en checklists waarderen, maar geholpen moeten worden bij het doorzien van de instrumenten. Dit vergt lestijd en toegepaste werkvormen. Studenten geven aan minder verschillende documenten te willen. In de toekomst zullen de checklists opgaan in de rubric, door uitbreiding van de rubric.

Terugblik op deelname aan de kenniskring van Monique Maas

Het verbeteren van de beroepspraktijk door praktijk-onderzoek is niet alleen een zaak van studenten, maar ook van docenten op lerarenopleidingen. Het systematisch doordenken van mogelijke oplossingen en verbeteringen geeft energie. Onderzoek doen is niet altijd makkelijk, goede onderzoeksvragen formuleren, goede methodieken doorgronden en leren toepassen en uiteindelijk verslaglegging van het onderzochte vergt tijd. Deelname aan een kenniskring helpt: samen stukken lezen en doorgronden, elkaars onderzoeken bespreken en van commentaar voorzien is leerzaam en nuttig. Voor mij heeft het lectoraat het doen van praktijkonderzoek mogelijk gemaakt - gerichte feedback en sturing op inhoud door sparren met collega's o.l.v. een betrokken lector is nuttig, leerzaam en plezierig. Het feit dat de gedane moeite ook echt vruchten afwerpt stimuleert enorm.



Hoofdstuk 6

De invloed van Random Assessment met Peer-Assessment (RAPA) op het studiegedrag van deeltijdstudenten wiskunde

Jos C.M. Verhoosel

6.1 Samenvatting

Van september 2010 tot februari 2011 is bij een groep van 56 deeltijdstudenten van de vakgroep wiskunde aan Fontys Lerarenopleiding in Tilburg een verkennend onderzoek gedaan naar peer assessment bij groepswork waarbij per groep één volkomen willekeurig gekozen (random) eindproduct door de docent beoordeeld werd, kortweg Random Assessment met Peer Assessment (RAPA) genoemd. De centrale onderzoeksvraag hierbij was: *Welk effect heeft random assessment met peerassessment op het studiegedrag van studenten?* Zowel aan het begin van de cursus als aan het einde is een quickscan afgenomen met betrekking tot studiegedrag en meeliftgedrag. Uit de analyse van deze scans kan de voorzichtige conclusie worden getrokken dat het studiegedrag door RAPA niet significant verandert. De manier waarop studenten met elkaar omgaan blijkt in positieve zin te worden beïnvloed. RAPA lijkt bij de beoordeling van grote hoeveelheden portfolio's de correctiedruk voor de docent te verlagen.

6.2 Inleiding

Docenten besteden zeer veel tijd aan toetsen (Plake, 1993) en vooral het nakijken van producten, zoals portfolio's neemt veel tijd in beslag. Een van de mogelijkheden om te komen tot verlaging van de correctietijd voor de docenten is het gebruik maken van peerassessment (Topping, 2009). De studenten worden hierbij medeverantwoordelijk voor het beoordelen van producten of processen van groepsgenoten. Verdere reductie van de correctietijd zou kunnen worden gerealiseerd als de docent niet alle producten van alle groepsleden hoeft te beoordelen maar per groepje slechts één (volstrekt random gekozen) product beoordeelt, met de opzet dat het cijfer dat hieruit voortvloeit voor het groepje gaat gelden als docentcijfer. Als blijkt dat het studiegedrag van studenten hierdoor niet in negatieve zin beïnvloed wordt, kan deze vorm van evalueren wellicht ook bij andere cursussen worden ingezet. De vraag is wat er gebeurt met het leren van studenten als zij voor hun eindresultaat mede-

afhankelijk worden van het werk van anderen. Gaan zij meer verantwoordelijkheid nemen voor de kwaliteit van de producten van andere groepsleden als zij niet op voorhand weten welke student uit de groep (als controle op het proces binnen de gehele groep) door de docent geselecteerd gaat worden? Groepsprocessen spelen hierbij duidelijk een rol. Daarnaast kan men zich afvragen in hoeverre het ethisch verantwoord is studenten een groepscijfer toe te kennen.

Studiegedrag en het werken in groepjes

Bij het werken in groepen spelen diverse groepsdynamische processen een rol. Een van de factoren hierbij is *sociale interdependentie*. Hiervan is sprake als de groepsleden gemeenschappelijke doelen hebben en de resultaten die groepsleden behalen beïnvloed worden door het doen en laten van de andere groepsleden. Johnson & Johnson (2005) geven aan dat competitiviteit, individualisme en de wil tot samenwerken hierbij sleutelfactoren zijn. Met betrekking tot groepsdoelen is er sprake van *positieve interactie* als de groepsleden ervan overtuigd zijn dat ze deze groepsdoelen alleen kunnen bereiken als ze onderling samenwerken en elkaar voortdurend stimuleren en faciliteren (Johnson, Johnson, & Smith, 2007).

Een andere factor is de mate waarin een groep zich een groep voelt: de *groepscohesie* genaamd. Die wordt door elke lid beïnvloed door de mate waarin competitiviteit, individualiteit of de wil tot samenwerking domineren. Groepscohesie wordt eveneens beïnvloed door enerzijds

de mate van openheid en delen (*vertrouwd gedrag*) en anderzijds acceptatie en steun (*betrouwbaar gedrag*). Daarnaast stelt groepswork ook eisen aan de onderlinge communicatie en het kunnen geven en ontvangen van feedback.

Bij groepswork kan sprake zijn van meeliftgedrag (Davies, 2009; Topping, 2009). Dit is een ongewenst verschijnsel bij groepswork dat gereduceerd en zo mogelijk voorkomen moet worden. Met betrekking hiertoe doen Karau en Williams (1993, 1995) een aantal aanbevelingen om dit te reduceren. In dat geval moet ervoor gezorgd worden dat:

1. De bijdragen van elk groepslid duidelijk herkenbaar zijn;
2. De doelen betekenisvol zijn, uitdagend, belangrijk en attractief zijn;
3. De prikkel om een goede prestatie te leveren sterk is;
4. De groepsleden herkennen en beseffen dat hun eigen inzet onmisbaar is voor een betere groepsprestatie;
5. Er sprake is van een grote groepscohesie;
6. De verrichtingen van elk groepslid zowel door de groepsleden als door de docent beoordeeld worden.

Het lijkt aannemelijk dat als er een groter beroep wordt gedaan op deze elementen, de mate waarin free-riding, meeliftgedrag, afneemt ook groter zal zijn.

Onderzoek door Maiden & Perry (2011) laat zien dat studenten groepswork hoog waarderen maar tevens dat elke methode die bijdraagt tot vermindering van 'free-riding' verwelkomd wordt.

Correctiedruk

Het ter controle beoordelen van slechts één product per groepje doet vermoeden dat de werkdruk van de docent hiermee zeker zal afnemen. Falchikov (2001) geeft echter aan dat het zo kan zijn dat hierbij op korte of middellange termijn wellicht weinig of geen tijdsbesparing voor de docent optreedt omdat er voor goed peerassessment ook tijd vrijgemaakt moet worden voor de voorbereiding, de organisatie, training en monitoring van studenten.

Peer assessment

Topping (2009) definieert peer assessment (vrij vertaald) als een opzet waarbij de lerende uitspraken moet doen over de specifieke niveaus en/of de waarde of kwaliteit van een product of gedragsindicatoren bij (gelijkwaardige) andere lerenden. Dat kan dan bijvoorbeeld gebeuren in de vorm van feedback, face-to-face of anders. Peer assessment kan toegepast worden op een groot aantal zaken waaronder schriftelijk werk, portfolio's, mondelinge presentaties, etcetera.

Assessment levert ook emotionele stress op. Lerenden hebben niet de neiging zich de positieve ervaringen met assessment te herinneren en zoeken gewoonlijk niet actief naar gelegenheden om zichzelf te assessen of ge-assessed te worden. Omdat peer assessment niet alleen op school plaatsvindt maar gedurende het hele leven zal naar verwachting iedereen van tijd tot tijd optreden als beoordelaar(assessor) of als beoordeelde (assessee). Het gebruik van peer assessment op school kan dan ook

vaardigheden opleveren die voor de rest van iemands leven bruikbaar zijn. (Topping, 2009; Boud & Falchikov, 2006).

Topping (2009) geeft hierbij een aantal aanbevelingen met betrekking tot de organisatie van peer assessments. Hij benoemt onder andere de volgende aspecten als belangrijk:

1. Het bespreken van de doelen de rationale, de verwachtingen en aanvaardbaarheid van de doelen. Deze dienen vooraf verhelderd te zijn. Dit geldt ook voor de aard en inhoud van het product dat beoordeeld gaat worden;
2. De criteria voor assessment in overleg en samen met de groep vooraf opstellen. Dit is belangrijk voor het eigenaarschap en het wegnemen van onzekerheden en angsten;
3. De samenstelling van de groepjes (bv. liever geen intieme vrienden of vijanden bij elkaar, mik op ongeveer dezelfde prestatieniveaus);
4. Het ervoor zorgen dat ze geregeld onderling overleg organiseren;
5. Het zorgen voor voldoende training, voorbeelden en praktijk. De kwaliteit van training maakt hierbij veel uit;
6. Het zo mogelijk verstrekken van richtlijnen en checklists;
7. Het bewaken van de voortgang en stel hierover eisen. Monitor en coachen zijn hierbij belangrijk;
8. Het controleren en bespreken van de kwaliteit van de feedback die studenten aan elkaar geven;

9. Het met regelmaat controleren of de eigen observaties overeenkomen met die van de peers van de studenten;
10. Het regelmatig praten met de studenten, evalueren en feedback geven aan studenten over dat wat je waarneemt over het proces en over de bijeenkomsten.

Topping (2009) merkt verder op dat peer-assessoren, ook als ze weinig ervaring hebben, assessments leveren met een minstens even grote betrouwbaarheid en validiteit als het assessment dat door de docent zelf wordt uitgevoerd mits ze ruim de tijd krijgen voor het uitvoeren hiervan. Daarnaast moet ook in gedachten worden gehouden dat studenten peerassessment minder prettig vinden omdat dit een grotere kans op kritiek van peers geeft (Ryan et al, 2007).

Het bovenstaande lijkt aannemelijk te maken dat een context waarin assessment een rol speelt alsook een grote afhankelijkheid van de kwaliteit van het werk van andere groepsleden, dit invloed zal hebben op het studiegedrag van studenten.

Op basis van het bovenstaande is voor het vak LOB gekozen voor de volgende werkvorm en manier van beoordelen:

- Studenten werken samen in groepjes van vier personen. De studenten stellen deze groepjes zelf samen op basis van de mededeling dat ze hierin gedurende een half jaar moeten samenwerken.
- De diverse thema's die bij LOB aan bod komen, worden aangeboden in vorm van instructies, lezingen, filmmateriaal, simulaties, discussies over thema's. Bij het woord thema moet in dit geval gedacht worden

aan bijvoorbeeld: (1) Wat is LOB? (2) de leerbiografie (3) breinleren (4) diversiteit en keuzeprocessen (5) kwaliteiten- en motievenreflectie (6) netwerken (7) het voeren van loopbaandialogen etc.

- Voor elk thema zijn een of twee studieopdrachten geformuleerd. Studenten werken deze opdrachten uit. Sommige opdrachten moeten individueel worden uitgewerkt, andere door de groep. Soms moeten studenten deze in hun eigen portfolio aanvullen met persoonlijke inzichten en standpunten.
- De portfolio's van de studenten (met daarin o.a. de uitwerkingen van deelopdrachten) rouleren tijdens de cursus binnen de eigen groep. De groepsleden toetsen de kwaliteit van de uitwerkingen aan de hand van een verstrekte rubric en geven elkaar feedback over het geleverde werk. Iedere student verbetert op basis van de verkregen feedback zijn werk. De rubric (zie bijlage 2) is vooraf doorgesproken met de studenten. De criteria die in de rubric gebruikt zijn, zijn echter niet door de studenten zelf opgesteld.
- Aan het einde van de cursus geven de groepsleden elke medestudent uit de groep een waardering voor het geproduceerde portfolio. Het gemiddelde cijfer hiervan geldt als peercijfer. Hierdoor wordt het voor studenten belangrijk dat ze ervoor zorgen dat het product van hun collega's ook echt goed is. Een slecht product betekent voor hem/haar immers een lager eindcijfer.
- Aan het einde van de cursus wordt *random* één van de portfolio's uit de groep beoordeeld door de begeleidende docent. Die gebruikt dezelfde rubric

als die de studenten hebben gehanteerd. Het hieruit voortvloeiende cijfer geldt voor elk groepslid als docentcijfer.

- Het eindcijfer wordt bepaald met de uitdrukking $(2 * \text{peercijfer} + \text{docentcijfer})/3$
- Bij twijfel, bijvoorbeeld met betrekking tot de wijze waarop studenten gefunctioneerd of beoordeeld hebben, kan de docent aanvullend een CGI (criterium gericht interview) afnemen.

Deze beoordelingsvorm duiden we verder aan met RAPA (Random selected Assessment with Peer Assessment).

6.3 Onderzoeksvragen

In dit onderzoek staat de volgende onderzoeksvraag centraal:

Welk effect heeft een assessmentvorm waarbij random een product uit een groep beoordeeld wordt (RAPA) op het studiegedrag van de studenten?

Om deze onderzoeksvraag te beantwoorden wordt het begrip studiegedrag verder geoperationaliseerd in de volgende deelvragen:

- In hoeverre verandert het vertrouwd gedrag en onbetrouwbaar gedrag van studenten door het gebruik van RAPA?
- In hoeverre verandert door RAPA de *interactie* tussen studenten?
- In hoeverre heeft het gebruik van RAPA invloed op de *sociale interdependentie* binnen een groep?

- In hoeverre heeft het gebruik van RAPA invloed op het *meeliftgedrag* van studenten?

De verwachting is dat de sociale interdependentie, het verantwoordelijkheidsgevoel voor de leerwinst van de ander en de mate van sturing om te komen tot een goed portfolio in positieve zin worden beïnvloed als gevolg van de beoordelingsvorm RAPA.

6.4 Methode

Participanten en context

Het onderzoek vond plaats in de eerste helft van het studiejaar 2010-2011. De studenten volgden het vak LOB (Loopbaan Oriëntatie en Begeleiding). Voor studenten aan de Fontys Lerarenopleiding in Tilburg is het vak LOB in het curriculum opgenomen. Het vak beoogt studenten vaardigheden en kennis bij te brengen waarmee ze in staat zijn om leerlingen te begeleiden bij keuzeprocessen. Op meta-niveau beoogt dit vak eveneens:

- het simuleren van samenwerkingsverbanden op de school waarin (vak)docenten samenwerken aan LOB (maar in het begin een verschillende visie hebben op LOB).
- het leren omgaan met de groepsdynamische processen die hiermee samengaan
- het 'elkaar beter willen maken', een toename dus van het verantwoordelijkheidsgevoel met betrekking tot het leren van de ander. (zie bv. modulewijzer LOB deeltijd wiskunde)

Voor de evaluatie van de studieprestaties van studenten wordt bij LOB gebruik gemaakt van portfolio's. Gebruikelijk is dat het bij dit vak om kleine groepen gaat (10 tot 20 studenten). Evaluatie van de studieprestaties gebeurt dan door middel van een individueel portfolio. Met ingang van 2010-2011 werd verwacht dat de aan LOB deelnemende groepen behoorlijk groter zouden worden (2 deeltijdgroepen van ongeveer 60 personen). Deze grootte zou tot gevolg kunnen hebben dat bij een ongewijzigde manier van werken de begeleidingstijd en correctietijd voor docenten onverantwoord groot zouden worden. Hierdoor ontstond de behoefte aan een werkvorm, die enerzijds aansloot bij de doelen van het vak en anderzijds zo mogelijk de werkdruk voor de docent zou beperken.

De bestudeerde populatie bestond uit 22 mannen en 34 vrouwen die een deeltijdstudie wiskunde volgden aan de Fontys Lerarenopleiding Tilburg. De gemiddelde leeftijd bedroeg 37 jaar met een standaarddeviatie van 9,2.

Instrumenten

Om te meten of RAPA effect heeft op de genoemde variabelen is een vragenlijst ontwikkeld op basis van bestaande vragenlijsten in het boek over groepsdynamica van Johnson en Johnson (2008). De ontwikkelde vragenlijst bestond uit 52 vragen op een vijf-punts-Likertschaal. (bijlage 6.1). De gebruikte schalen zijn *vertrouwd- en betrouwbaar gedrag* (deelvraag a), *positieve interactie* (deelvraag b), *reductie van meeliftgedrag* (deelvraag d), *groepsdruk*, *kritisch samenwerken* en *intrinsieke motivatie*

en *sociale interdependentie* bestaande uit *competitiviteit*, *individualisme* en *coöperativiteit* (deelvraag c).

Een betrouwbaarheidsmeting voor de vragenlijst van 52 items leverde een Cronbach's Alpha van 0.826.

In Tabel 6.1 is weergegeven hoe de vragen uit de vragenlijst zijn onder te brengen in (de eerder genoemde) schalen.

Tabel 6.1 Relatie tussen de vragen uit de vragenlijst en de schalen

Schaal	Vragen uit de vragenlijst die hieraan gerelateerd zijn
Vertrouwd gedrag	10A, 8Br, 8D, 8E, 8F
Betrouwbaar gedrag	10C, 10G, 10H
Positieve interactie	9A, 10B, 10C, 10D, 10E, 10F, 10G, 10H, 10Ir, 10J, 10K, 10L, 10M, 10N
Groepsdruk	11A, 11H
Kritisch samenwerken	11B, 11F, 11Ir, 11J, 11K
Intrinsieke motivatie	11C, 11D, 11E, 11G
Competitief gedrag	8Ar, 8F, 8I
Individualisme	8B, 8D, 8G, 8J
Coöperativiteit	8C, 8E, 8H, 8K

Noot: Achter sommige vraagnummers staat een r. Dit geeft aan dat de vraag gehercodeerd is.

Gesprekken, cursus-evaluaties en reflecties van studenten (kwalitatief)

Na afloop van de cursus worden opmerkingen uit gesprekken, cursus-evaluaties en reflecties van studenten

geanalyseerd. Ze worden ingedeeld in de volgende categorieën:

1. Opmerkingen over hoe de cursus hun denken beïnvloed heeft;
2. Opmerkingen over belemmeringen van organisatorische aard;
3. Opmerkingen over belemmeringen die te maken hadden met de persoonlijke ontwikkeling van de student;
4. Belemmeringen die de groepsvoortgang bemoeilijkten;
5. Belemmeringen die voortvloeien uit een onvoldoende voorbereiding van studenten in het voortraject.

Deze opmerkingen worden gebruikt om zicht te krijgen op niet voorziene factoren.

Data-analyse

De verandering van het studiegedrag werd bepaald door te kijken naar het verschil tussen de gemiddelde score aan het begin van de cursus en die aan het einde van de cursus zoals die behaald konden worden voor bepaalde schaalvariabelen. In het begin van de cursus hebben studenten zelf de groepjes samengesteld. Dit o.a. op basis van criteria als bv. "samen reizen". Hierdoor was het niet goed mogelijk om te kunnen garanderen dat een controlegroep gelijkwaardig zou zijn aan de experimentele groep. Vanwege deze reden is afgezien van het gebruik van een controlegroep.

Er werd gebruik gemaakt van een pre- en post-test die afgenomen werd tijdens de eerste (N=56) respectievelijk de laatste bijeenkomst (N=54). Eerst is gekeken of op itemniveau verschillen te zien waren tussen de voor- en nameting. Daarna is gekeken hoe er op de genoemde schalen gescoord is (rechte tellingen met gemiddelde en sd) en of er significante verschillen gevonden zijn tussen de voor- en de nameting.

Vervolgens is gekeken of de schalen zoals ze vooraf geconstrueerd waren ook bij deze groep op dezelfde manier te herleiden zijn. Hiervoor is een factoranalyse met principale componenten uitgevoerd om het aantal factoren binnen een groep vragen te reduceren. Omdat bleek dat er andere schalen af te leiden waren, is gekeken of er op deze nieuwe schalen effecten waarneembaar zijn tussen de pre- en postmetingen. Sommige vraagitems bleken daarbij met geen enkel ander vraagitem te correleren. Deze zijn in de latere berekeningen daarom niet meer meegenomen.

6.5 Resultaten

In Tabel 6.2 staan de gemiddelde scores en standaard deviaties van die afzonderlijke items die een significante toename of afname lieten zien.

Tabel 6.2 Gemiddelde scores, standaarddeviaties van de voormeting en de nameting en resultaat *t*-toets.

Item	Voormeting		Nameting				
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>T</i>	<i>df</i>	<i>p</i>
(3) Adviseert medestudenten procesmatig bij leertaken	2.27	0.70	2.61	0.76	-2.22	92	0.029
(8J) Binnen de groep werken we liever alleen dan met elkaar	2.20	0.91	2.62	0.92	-2.14	92	0.035
(9B) We voelen ons persoonlijk betrokken bij de prestaties van de groep	4.14	0.62	3.82	0.77	2.28	92	0.024
(10B) Tijdens groepsgesprekken houd ik mijn gevoelens, reacties, ideeën en reacties voor mijzelf	2.23	0.76	1.87	0.70	2.34	92	0.021
(10G) Ik accepteer en steun de openheid van andere groepsleden.	4.14	0.62	3.82	0.77	-1.99	92	0.050

Op item-niveau is een significante *toename* gevonden voor de voorkeur om binnen de groep alleen te werken. Dit gold ook voor de mate waarmee ze elkaar procesmatig adviseerde bij leertaken. Daarnaast is een significante *afname* te zien van de persoonlijke betrokkenheid en prestaties van de groep als geheel, het accepteren en steunen van de openheid van de andere groepsleden en het voor zichzelf houden van gevoelens, ideeën en reacties.

Tabel 6.3 Gemiddelde scores en standaard deviaties op de afzonderlijke schalen en de gemeten verschillen.

Item	Voormeting		Nameting				
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>p</i>
Vertrouwd gedrag	11,48	1,32	11,86	1,61	-0,38	88	0,220
Betrouwbaar gedrag	19,36	2,23	20,19	2,09	-0,83	88	0,078
Positieve interactie	45,54	4,19	44,51	5,58	1,02	87	0,325
Groepsdruk	7,47	1,12	7,58	1,39	-0,11	89	0,685
Kritisch samenwerken	17,44	2,81	18,00	3,37	-0,56	90	0,386
Intrinsieke motivatie	14,04	2,78	14,00	2,28	0,04	92	0,948
Competitief gedrag	38,94	5,67	39,51	6,07	-0,6	87	0,650
Individualisme	7,385	1,35	7,05	1,06	0,33	88	0,211
Coöperativiteit	10,47	2,49	11,43	2,42	-0,96	90	0,070

Op schaalniveau zijn geen significante verschillen waarneembaar tussen de voor- en nametingen voor de diverse schalen.

Gesprekken met studenten (kwalitatief)

Uit de reacties van studenten is met betrekking tot de 5 categorieën het volgende op te merken.

1. Opmerkingen over hoe de cursus hun denken heeft beïnvloed
 - a. “Ik praat nu heel anders met de kinderen uit mijn eigen gezin!”
 - b. Ontwikkeling van de eigen visie op LOB
2. Opmerkingen over belemmeringen van organisatorische aard
 - a. *Grote reisafstanden.*
Het fysiek bij elkaar komen voor overleg werd vaak bemoeilijkt door het gegeven dat veel studenten ver weg woonden van elkaar
 - b. *Beschikbaarheid van een stageschool.*
Binnen de groepen hadden niet alle studenten een stageschool waarop de stage gebonden opdrachten konden worden uitgevoerd. Dit werkte vertragend op de voortgang. Sommige groepjes konden hierdoor pas veel later hun portfolio inleveren
 - c. *ICT-vaardigheid m.b.t. overleg op afstand.*
Studenten kunnen vaak wel met email omgaan maar zijn onbekend met vormen van afstandsonderwijs. Kennis hierover kan bevorderend zijn voor het groepsproces.
3. Opmerkingen over belemmeringen die te maken hadden met de persoonlijke ontwikkeling van student
 - a. *Druk vanwege de dagelijkse praktijk.*
Enkele studenten gaven aan dat hun baan vaak prioriteit had boven de cursus. Dit leidde dan vaak tot uitstelgedrag of tot een geringere inzet.
 - b. *Zelfdiscipline van studenten*
4. Belemmeringen die de groepsvoortgang bemoeilijkte
 - a. *Samenstelling van de groepjes.*
Sommige groepjes zijn geboren uit “nood” (te laat binnengekomen, overgebleven etc.). Dit is de samenwerking binnen de groepjes soms opgebroken. In voorkomende gevallen werd hierbij de groepssamenstelling aangepast.
 - b. *Uitstelgedrag.*
Soms belemmerde het uitstelgedrag van individuele studenten de voortgang van de groep
 - c. *Sociale vaardigheden.*
Binnen sommige groepen ontstonden problemen die niet goed werden opgelost. Er waren aanwijzingen dat dit te maken had met onvoldoende ontwikkelde communicatieve vaardigheden.
5. Belemmeringen die voortvloeien uit een onvoldoende voorbereiding van studenten in het voortraject
 - a. *Vaardigheden m.b.t. het geven van constructieve feedback.*

6.6 Conclusie en discussie

Met betrekking tot de onderzoeksvragen zijn op basis van deze studie de volgende conclusies te trekken.

Welk effect heeft RAPA op het studiegedrag van studenten?

Uit Tabel 6.3 blijkt dat RAPA, voor de beschouwde groep studenten, op geen van de gehanteerde schalen een significante invloed heeft gehad. Het studiegedrag lijkt in die zin niet meetbaar negatief of positief beïnvloed.

Op item-niveau zijn echter wél significante verschillen gemeten (zie Tabel 6.2). Hieruit zou men kunnen opmaken dat men zich binnen de groepjes enerzijds gemiddeld gezien vrijer is gaan voelen om dingen te uiten naar de ander (veiligheid) maar anderzijds ook minder accepteert dat anderen dit ook doet. Dit kan een aanwijzing zijn om meer aandacht te geven aan de vaardigheid van het geven en ontvangen van feedback. De toegenomen voorkeur voor het alleen werken en de afgenomen betrokkenheid bij de groep lijkt hiermee causaal samen te hangen.

Meta-analyse: nieuwe schalen

Omdat er geen significante verschillen te zien waren voor de gebruikte schalen is vervolgens nagegaan of de schalen zoals ze vooraf geconstrueerd waren ook bij deze groep op de zelfde manier te herleiden zijn. Hiervoor is een factoranalyse met principale componenten uitgevoerd om het aantal factoren binnen een groep vragen te reduceren. Omdat bleek dat er andere schalen af te leiden waren, is gekeken of er op deze nieuwe

schalen effecten waarneembaar zijn tussen de pre- en postmetingen. Sommige items bleken daarbij met geen enkel ander vraagitem te correleren. Deze zijn in de latere berekeningen daarom niet meer meegenomen.

De analyse met betrekking tot deze nieuwe schalen bleek eveneens geen significante verschillen aan het licht te brengen met uitzondering van de nieuwe schaal “stress binnen de groep”. Deze lijkt significant af te nemen ($t = 1,986$; $df = 92$; $p = 0,05$).

Samenvattend kan men voorzichtig concluderen dat RAPA voor deze groep deeltijdstudenten geen meetbare verandering teweeg heeft gebracht met betrekking tot de studiehouding.

Reductie van de begeleidings- en correctietijd

Kijkend naar het proces als geheel merkt Falchikov (2001) op dat goed peer-assessment binnen de groepen een beslag legt op tijd, organiseren en monitoring. Deze aspecten zijn vanwege het experimentele karakter van deze studie niet volledig gemonitord. Hiervoor liepen de processen van begeleiding en beoordeling te dicht naast elkaar. De tijd, nodig voor het beoordelen van alle portfolio's, werd bepaald door diverse factoren: (1) de tijd die nodig is voor het instrueren van de groepen voor de nieuwe werkwijze met daarbij komende mogelijk tijdrovende extra activiteiten (2) de tijd die nodig is voor het beoordelen van één portfolio van een groep (3) het berekenen van de individuele cijfers (4) de nabespreking met individuele studenten en groepjes. Hoewel de

correctiedruk *ervaren* is als minder belastend, is voor harde uitspraken met betrekking tot de invloed van RAPA op de begeleidings- en correctietijd aanvullend onderzoek nodig.

Eerlijke beoordeling

In dit experiment hebben studenten zelf hun groepjes samengesteld. Daarbij doet zich de mogelijkheid voor dat er zwakkere groepsleden bij zitten. Hoe eerlijk is het dan dat die door het groepsgebeuren hoger scoren dan ze waard zijn? Nordberg (2008) geeft aan dat er bij groepswork ongetwijfeld meer geleerd wordt maar dat de beoordeling vermoedelijk ook minder eerlijk is.

De invloed van de weegfactoren is vooraf met Excel gesimuleerd. Uitgangspunt hierbij was dat sterk presterende studenten een hogere eindscore moesten kunnen behalen dan de zwakker presterende studenten. De weegfactoren zijn hierbij tevens zo gekozen dat de verschillen niet te groot zouden worden.

Aanvullend onderzoek

In het cursusjaar 2012-2013 is een variant van RAPA uitgetest waarbij de portfolio's weliswaar rouleerden tijdens de cursus maar "het elkaar een cijfer geven voor het opgestelde portfolio" was weggelaten. Hierbij bleek dat "het elkaar beoordelen" een essentieel onderdeel is voor RAPA. Het achterwege laten hiervan gaf studenten meer gelegenheid zich te onttrekken aan een serieuze discussie over de kwaliteit van het door de ander geleverde werk.

6.7 Aanbevelingen

1. Uit de studie is niet gebleken dat RAPA invloed heeft op het studiegedrag van studenten. Tegelijkertijd kan worden opgemerkt dat RAPA wel geschikt lijkt te zijn voor het verlagen van de correctiedruk bij het gebruik van grote hoeveelheden portfolio's. Voor het kunnen doen van harde uitspraken hierover is aanvullend onderzoek nodig.
2. Het elkaar een peercijfer toekennen is essentieel voor RAPA. Studenten hebben hierdoor geen gelegenheid zich te onttrekken aan een serieuze discussie over de kwaliteit van het door de ander geleverde werk.
3. Tijdens de uitvoering van het experiment is geen speciale aandacht gegeven aan de manier waarop de groepen intern aan elkaar feedback gaven. Studenten gaven in hun reflecties, evaluaties en feedback na afloop van de cursus aan dat hier meer aandacht voor zou moeten zijn. Volgens Topping (2009) en Davies (2009) heeft extra voorbereiding van de groepen met betrekking tot het geven van constructieve feedback en het monitoren/coachen een positief effect op het groepsproces. Uit onderzoek van Xiao (2010) en Topping (2009) blijkt verder dat peer-assessment-training een gunstige invloed heeft op de kwaliteit van hun schrijven, op de schriftelijke feedback die studenten aan elkaar geven en op de betrouwbaarheid van de assessment-

scores. Ook blijkt dat er dan een hogere tevredenheid optreedt met betrekking tot peer-feedback. Het verdient dan ook aanbeveling om peer-assessment-training expliciet op te nemen in het voortraject van RAPA.

4. Tijdens de cursus bleek dat er een groot beroep gedaan werd op de vaardigheden van de begeleiders met betrekking tot de coaching van groepsprocessen. Dit is een punt van aandacht.

Bijlage 6.1 Vragenlijst LOB-studenten deeltijd

Beste student,
Zoals aangekondigd proberen we de ontwikkeling van de activiteiten binnen de studiegroepen te monitoren. In dit kader is een vragenlijst opgesteld die we een keer na ongeveer 2 bijeenkomsten en een keer na afloop afnemen. Uit de vergelijking van deze resultaten hopen we conclusies te kunnen trekken met betrekking tot het effect van de vorm en werkwijze bij deze manier van werken.

Het is de bedoeling dat de vragen individueel beantwoord worden.

In verband met de privacy zullen de antwoorden volstrekt vertrouwelijk worden gehanteerd en na afloop volledig geanonimiseerd worden. De antwoorden hebben geen invloed op de wijze waarop je eindcijfer voor LOB tot stand komt.

Algemene gegevens

Naam:

Leeftijd:

Geslacht:* M / V

Geeft momenteel ook les:* J / N

Leservaring:* nee / een beetje / veel

* omcirkelen wat van toepassing is

Bij de vragen 1 t/m 7, het bolletje dat van toepassing is zwart maken

Vraag 1. In hoeverre voel jij je verantwoordelijk voor de ontwikkeling van je medestudenten?

- ☐ niet
- ☐ een beetje
- ☐ zeer

Vraag 2. In hoeverre adviseer je je medestudenten *inhoudelijk* bij het schrijven/uitwerken van opdrachten en/of leertaken?

- ☐ niet
- ☐ alleen als ze er om vragen
- ☐ af en toe (op eigen initiatief)
- ☐ regelmatig (op eigen initiatief)

Vraag 3. In hoeverre adviseer je je medestudenten *procesmatig* bij het schrijven/uitwerken van opdrachten en/of leertaken?

- ☐ niet
- ☐ alleen als ze er om vragen
- ☐ af en toe (op eigen initiatief)
- ☐ regelmatig (op eigen initiatief)

Vraag 4. Met welk *doel* zou je je medestudenten adviseren bij het schrijven/uitwerken van opdrachten en/of leertaken? (meerdere antwoorden mogelijk!!!)

- ☐ om er zelf beter van te worden
- ☐ om de medestudent te helpen bij zijn/haar ontwikkeling
- ☐ om in de toekomst er samen beter van te worden

Vraag 5. Als je moet werken aan een portfolio-opdracht/leertaak, in hoeverre verwacht je dan advies van je medestudenten?

- ☐ niet
- ☐ alleen als ik er om vraag
- ☐ af en toe, zonder dat ik er om vraag
- ☐ regelmatig, zonder dat ik er om vraag

Vraag 6. Hoe belangrijk vind je het om samen te werken met medestudenten om te komen tot een individueel eindproduct?

- ☐ niet
- ☐ een beetje
- ☐ zeer

Vraag 7. Hoe belangrijk vind je het om samen te werken met medestudenten met betrekking tot je eigen ontwikkeling?

- ☐ niet
- ☐ een beetje
- ☐ zeer

Vraag 8. Geef bij de volgende uitspraken aan in hoeverre deze globaal genomen opgaan.

Vul het cijfer in dat jouw gedrag het beste weergeeft:

- 1 = nooit 2 = zelden 3 = soms
4 = meestal 5 = altijd

Nr.	Uitspraak	Cijfer
A	Ik vergelijk mezelf steeds met anderen uit de groep om te zien wie het beste is.	
B	Binnen de groep werken we een groot deel van de tijd zelfstandig	
C	Binnen de groep worden ideeën en middelen volop gedeeld	
D	Binnen de groep wordt graag zelfstandig gewerkt.	

E	We leren veel belangrijke dingen van elkaar	
F	Binnen de groep bestaat de neiging om elkaar te overtroeven	
G	Samenwerken vinden we best lastig	
H	We helpen elkaar om de zaakjes goed af te krijgen	
I	Binnen de groep stimuleren we elkaar om beter te zijn dan een ander	
J	Binnen de groep werken we liever alleen dan met elkaar	
K	Binnen de groep heerst de overtuiging dat we productiever zijn als we met elkaar samenwerken	
L	Binnen de groep is sprake van positieve interactie	

Vraag 9. Geef bij de volgende uitspraken aan in hoeverre ze binnen de groep, naar jouw inzicht, globaal genomen opgaat. Vul het cijfer in dat jouw gedrag het beste weergeeft:

1 = nooit 2 = zelden 3 = soms
4 = meestal 5 = altijd

Nr.	Uitspraak	Cijfer
A	We helpen elkaar op een effectieve en efficiënte manier	
B	We voelen ons persoonlijk betrokken bij de prestaties van de groep	
C	We luisteren kritisch naar elkaar	

D	We hebben oog voor elkaars activiteiten, geven daar onmiddellijk feedback op en verlenen elkaar zo nodig bijstand	
E	We zetten regelmatig vraagtekens achter de conclusies en argumenten van de ander. Op deze manier prikkelen we elkaars nieuwsgierigheid en motiveren we elkaar tot leren	
F	We stimuleren elkaar voor het realiseren van een gemeenschappelijk doel	
G	We vertrouwen elkaar, en zijn te vertrouwen	
H	Als iemand een betere manier heeft om een taak tot een goed einde te brengen dan nemen anderen die meestal snel over	
I	Binnen de groep ervaren we weinig angst en stress	
J	We besteden voortdurend aandacht aan de onderlinge samenwerking	
K	We stimuleren elkaar om onderling samen te werken	
L	Als één van ons faalt, voelt dat aan alsof wij als groep ook falen.	
M	Het resultaat van iemands werk wordt steeds goed bekeken en doorgesproken.	
N	Binnen de groep heb ik neiging het initiatief te nemen	

Vraag 10. Geef bij de volgende uitspraken aan in hoeverre ze binnen de groep, naar jouw inzicht, globaal genomen opgaat. Vul het cijfer in dat jouw gedrag het beste weergeeft:

1 = nooit 2 = zelden 3 = soms
4 = meestal 5 = altijd

Nr.	Uitspraak	Cijfer
A	Ik ben open en eerlijk in mijn contacten met de gehele groep	
B	Tijdens groeps gesprekken houd ik mijn gevoelens, reacties, ideeën en reacties voor mijzelf	
C	Ik laat de andere groepsleden weten dat ik me bewust ben en waardering heb voor hun capaciteiten, talenten en ervaringen	
D	Ik bied anderen mijn hulp aan om het prestatieniveau te verhogen	
E	Ik deel materialen, informatiebronnen of andere hulpmiddelen.	
F	Als er me iets dwarszit, bespreek ik dat met de groepsleden	
G	Ik accepteer en steun de openheid van andere groepsleden.	
H	Ik beoordeel de bijdrage van anderen in termen van het nut dat ze voor mij hebben en in termen van of zij het bij het rechte eind hebben of niet.	

Vraag 11. Geef bij de volgende uitspraken aan in hoeverre ze binnen de groep, naar jouw inzicht, globaal genomen opgaat. Vul het cijfer in dat jouw gedrag het beste weergeeft:

1 = nooit 2 = zelden 3 = soms
4 = meestal 5 = altijd

Nr.	Uitspraak	Cijfer
A	De bijdrage van iemand is goed herkenbaar voor iedereen	
B	De doelen binnen de groep hebben een duidelijk herkenbaar nut en persoonlijke betekenis voor ons toekomstig werk.	
C	De onderwerpen die binnen de groep aan bod komen vind ik belangrijk.	
D	De onderwerpen die binnen de groep aan bod komen vind ik uitdagend	
E	De onderwerpen die binnen de groep aan bod komen vind ik interessant	
F	Binnen de groep is een sterke prikkel om een goede prestatie te leveren	
G	Voor iedereen is duidelijk dat een hogere eigen inzet leidt tot een betere groepsprestatie	
H	Er is sprake van veel cohesie binnen de groep	
I	Binnen de groep beoordelen we voortdurend het werk van de ander	
J	Binnen de groep geeft iedereen aan de andere groepsleden feedback	
K	Binnen de groep ontvangt elk lid feedback van de andere groepsleden	

Bijlage 6.2 De rubric peer-assessment LOB*

	beginnend	in ontwikkeling	Competent	voorbeeldig
Aanwezigheid	Is aanwezig bij de bijeenkomsten. Geeft af en toe een bijdrage. ☐	Is aanwezig bij alle bijeenkomsten en geeft er in de bijeenkomsten duidelijk blijk van zich verdiept te hebben in relevante documenten. De bijdragen sluiten hierbij aan. ☐	...en luistert en reageert inhoudelijk en op betrekkningsniveau naar andere groepsleden ☐	...en neemt initiatieven voor nieuwe acties die aanleiding geven tot het vergroten van de eigen loopbaancompetenties en die van andere groepsleden. Er is sprake van sturing en coaching van anderen. ☐
groei eigen LOB-competenties	Beschrijft globaal de ontwikkeling van de eigen LOB-competenties. ☐	Koppelt de eigen ontwikkeling van LOB-competenties aan acties. ☐	...en er is sprake van een planmatige aanpak...en... organiseert feedback op eigen ontwikkeling van competenties. ☐	Is rolmodel voor medestudenten. ☐
Vormgeving portfolio	Het portfolio is een losse map met al dan niet volledig uitgewerkte portfolio-opdrachten. ☐	Het portfolio bestaat uit een map met uitgewerkte portfolio-opdrachten. Er is sprake van een logische indeling. Een inhoudsopgave aanwezig en ook een blz. nummering.... ☐	en is geïllustreerd met afbeeldingen... en overzichtelijk en aantrekkelijk vormgegeven ☐	... en is lay-out-technisch <i>functioneel</i> vormgegeven. ☐
Volledigheid	Het portfolio is voorzien van titelblad met studentgegevens ☐	Er ontbreken sommige of meerdere onderdelen zoals bijvoorbeeld: • een sterkte/zwakte-analyse m.b.t. de 5 loopbaancompetenties • opdrachten die uitgewerkt zouden moeten zijn • Een beschouwing over de eigen groei m.b.t. de 5 loopbaancompetenties • Een beschouwing over het groepsgebeuren ☐	Het dossier is compleet: titelblad, inhoudsopgave, volledig uitgewerkte opdrachten, S/Z-analyse, formulering van de van de eigen rol bij en visie op LOB, ☐	...en een terugblik op de cursus als geheel (in het kader van de eigen groei) inclusief suggesties voor aanpassing/verbetering ☐

	beginnend	in ontwikkeling	Competent	voorbeeldig
Kwaliteit van de uitgewerkte opdrachten	De uitwerkingen zijn kort en oppervlakkig. ☐	In de opdrachten wordt een koppeling gelegd met de eigen ervaringen, die van anderen en de eigen toekomstige rol op het gebied van LOB... ☐	en ...de student formuleert leervragen voor zichzelf ☐	... met aanzetten voor het integreren van de kennis over LOB in het eigen handelen ☐
De Eigen groei	Beschrijft hoe hij hieraan heeft gewerkt, ☐	...en betreft hierin de eigen vaardigheden en kennis met betrekking tot de vijf loopbaancompetenties: kwaliteitenreflectie, motievenreflectie, opleidingen- en werkexploratie, loopbaansturing en netwerken ☐	... en kijkt hierbij ook kritisch naar zichzelf ... en formuleert op dit gebied nieuwe leervragen voor zichzelf ☐	...en zet deze leervragen systematisch om in acties m.b.t. het eigen handelen ☐
Groepsgebeuren	Werkt voornamelijk voor zichzelf. Ervaringen vanuit de groep spelen geen (belangrijke) rol. Uitwisselen van ervaringen op het gebied van LOB n.a.v. de stage-ervaringen ☐	Er is sprake van een actieve opstelling. Neemt verantwoordelijkheid naar de groep toe en komt afspraken bv. m.b.t. de werkverdeling binnen de groep na. ☐	...en stuurt anderen op het gebied van kwaliteit en geeft constructieve feedback m.b.t. het functioneren van anderen binnen de groep ☐	...en bewaakt een positief groepsgebeuren door actief en bewust in te spelen op groepsdynamische processen ☐

Opbrengsten van het doen van onderzoek

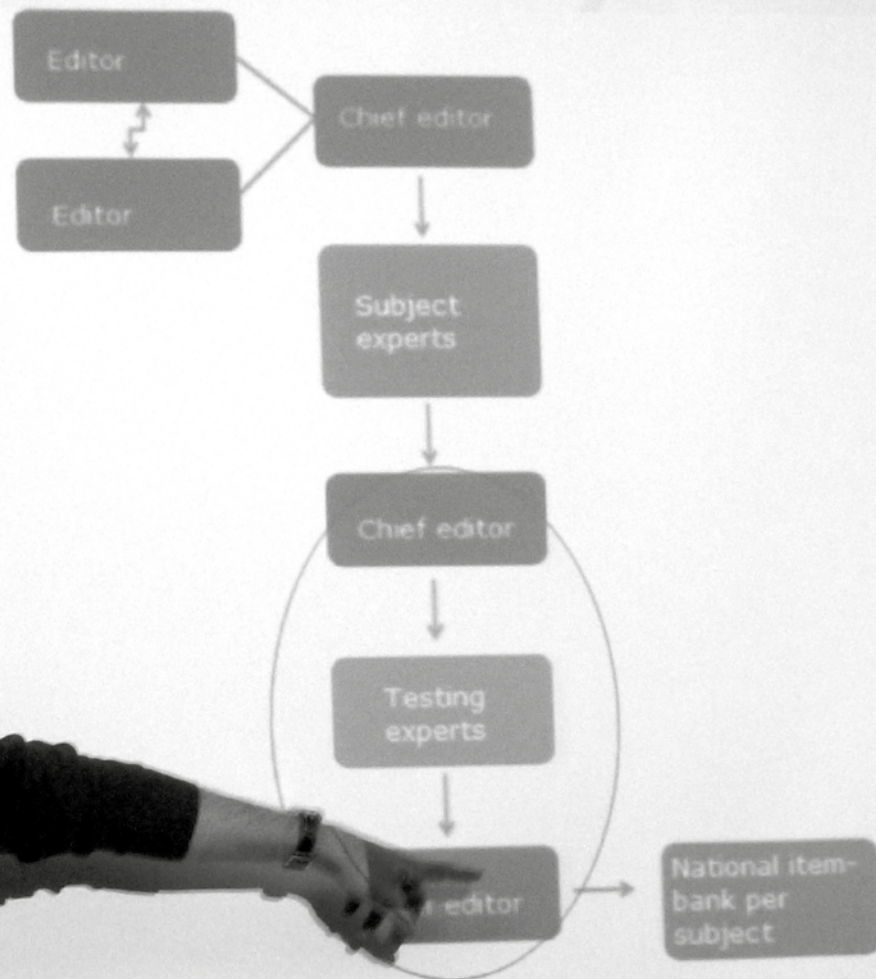
Jos Verhoosel

Het doen van onderzoek heeft mij persoonlijk meer inzicht gegeven in de aard van de problematiek waarmee (master) studenten tijdens het doen van onderzoek te maken kunnen krijgen. Denk hierbij aan problemen die zich kunnen voordoen met het verzamelen van meetgegevens, het verwerken van data, de druk van de dagelijkse bezigheden die ervoor kan zorgen dat de continuïteit van het onderzoek verstoord raakt, etc.

Voor mijzelf leverde dat niet alleen een meer kritische kijk op het onderwijs dat ik zelf geef maar ook een vergroting van de eigen vaardigheden met betrekking tot onderzoek doen alsook een meer kritische blik op onderzoek zoals dat door anderen wordt uitgevoerd. Ook de drempel met betrekking tot het onderzoeken van andere zaken uit de eigen onderwijspraktijk is verlaagd.



1.2 The context, items being made



Hoofdstuk 7

De landelijke kennistoets Geschiedenis

7.1 Vooraf

Het project *10voordeLeraar* is een landelijk project van de Vereniging Hogescholen met als doel totstandkoming en invoering van landelijke kennisbases, kennistoetsen en kennisbanken voor de bachelor- en masteropleidingen in het hbo die studenten opleiden tot leraar in het voortgezet onderwijs, het beroepsonderwijs of de volwasseneneducatie. Het lectoraat Eigentijds toetsen en beoordelen heeft een bijdrage aan dit project geleverd door het opzetten van een kwaliteitscommissie die de kwaliteit van de toetsen moest bevorderen. Dit gebeurde door workshops over item- en toetsconstructie in de ontwerpfase en door toets- en itemanalyse na afname van de pilottoetsen. Over deze werkzaamheden is samen met Arnoud van Leuven (op dat moment projectleider bij het project) en Dominique Sluijsmans (Hogeschool Zuyd) een artikel geschreven. Dit artikel is in 2012 gepubliceerd in *Kleio* en is de basis voor dit hoofdstuk.

Joosten-ten Brinke, D., Lansu, W., Van Leuven, A., & Sluijsmans, D. (2012). Gewogen en ...? De landelijke kennistoets geschiedenis. *KLEIO*, 53(1), 32-34. Het artikel beschrijft de werkwijze vanuit het vak Geschiedenis. Deze werkwijze is exemplarisch voor de werkwijze van alle vakken.

7.2 Samenvatting

Politiek en samenleving roepen om meer basiskennis bij docenten. Een mogelijk antwoord hierop is de kennisbasis voor lerarenopleidingen en een toets ter controle. Maar wat is een kennisbasistoets geschiedenis? En hoe komt de toets tot stand?

In het kader van de Kwaliteitsagenda Krachtig Meesterschap heeft de HBO-raad het project '*10voordeleraar*' (www.10voordeleraar.nl) opgezet met als doel de gezamenlijke kennisbases voor de lerarenopleidingen te beschrijven. Door het vaststellen van deze kennisbases wordt ernaar gestreefd de kennis en vaardigheden van de studenten in alle hbo-lerarenopleidingen te vergroten en de kwaliteit van de opleidingen te borgen, mede door de ontwikkeling van landelijke digitale toetsen. In 2010 is gestart met het ontwikkelen van de landelijke digitale kennistoetsen voor de tweedegraads opleidingen. In dit hoofdstuk worden de resultaten van en ervaringen met het ontwikkelen van de landelijke kennistoets voor geschiedenis gepresenteerd. Uit de resultaten blijkt dat voor een duurzame kwaliteit van de toets structurele ondersteuning bij de itemconstructie essentieel is. Tevens is een goede aansluiting met de kennisbasis in het curriculum voorwaarde om de toets optimaal in te zetten.

7.3 Wie is betrokken bij een landelijke toets?

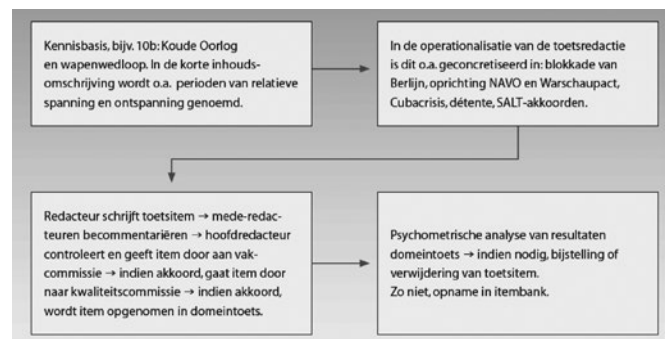
Het project '10voordeleraar' is een antwoord op de zorg in de samenleving en de politiek over het gebrek aan vakkennis van docenten in het basis- en voortgezet onderwijs (Inspectie van het Onderwijs, 2004; NVAO, 2004). Op basis van deze zorg heeft de minister aangegeven dat het eindniveau van hbo-studenten transparant en onderling vergelijkbaar moet worden vastgesteld (Ministerie van OC&W, 2008). Om dit te realiseren werken in opdracht van de overheid en onder regie van de HBO-raad (nu Vereniging Hogescholen) redactieteams van hogeschooldocenten aan landelijke digitale kennistoetsen waarmee het vakinhoudelijke eindniveau verhoogd en geborgd wordt. Vanaf 2013 krijgen alle studenten voor het vak geschiedenis een twee uur durende toets over de kernconcepten van dit vak. Pas na het slagen voor deze toets kan de student deelnemen aan het eind-assessment van de eigen opleiding.

Een belangrijk onderdeel bij de ontwikkeling van deze landelijke toets is de kwaliteitsborging en externe legitimering. Bij de toetsontwikkeling worden redactieteams kritisch gevolgd door onafhankelijke vakcommissies. Deze vakcommissies worden voorgezeten door een vertegenwoordiger uit het Landelijk Expertisecentrum Mens en Maatschappijvakken (www.expertisecentrum-mmv.nl). Toetsdeskundigen die lid zijn van de voor het project ingestelde kwaliteitscommissie beoordelen de items en de toets als geheel voor en na afname van de toets op toetstechnische aspecten.

Deze werkwijze wordt tot 2013 op projectmatige basis voortgezet. De projectleiders zijn verantwoordelijk voor het aansturen van het project en er is een adviesraad, bestaande uit hoogleraren op het gebied van toetsen en beoordelen, die het geheel monitort.

7.4 Van kennisbasis naar kennistoets

De kennisbasis geschiedenis (www.kennisbasis.nl) beschrijft de kennis die startende tweedegraads leraren geschiedenis minimaal moeten bezitten. Van de kennisbasis zijn twee domeinen aangewezen voor de landelijke toetsing, namelijk de tijdvakken en historisch redeneren. De toetsredactie heeft deze domeinen nader geconcretiseerd in gebeurtenissen, personen, ontwikkelingen en verschijnselen. Op basis van deze operationalisatie maken de redacteurs toetsitems. Dit redactieproces is in Figuur 7.1 schematisch weergegeven.



Figuur 7.1. Schematische weergave van het itemconstructieproces.

Op basis van de operationalisatie van de kennisbasis is een toetsmatrijs ontwikkeld waarin de inhoud en het gewenste beheersingsniveau tegen elkaar zijn afgezet. Kennisvragen vergen reproductie van kennis, terwijl toepassingsvragen van studenten verwachten dat ze op basis van nieuwe informatie vraagstukken kunnen oplossen. Hiervoor worden uitsluitend meerkeuzevragen gebruikt. Figuur 7.2 laat een deel van deze toetsmatrijs zien. Er is een vaste verdeling van de in de toets gestelde vragen over de verschillende domeinen. Deze verdeling komt overeen met de bandbreedte van de in de toetsmatrijs vermelde percentages. Iedere student krijgt dezelfde vragen, deze vragen worden per blok echter gerandomiseerd uitgeserveerd.

Tijdvak 1: Jagers en boeren	40-70%	10-30%	20-50%
Tijdvak 2: Grieken en Romeinen	40-70%	10-30%	20-50%
Historisch redeneren tijdvak 2			100%
Tijdvak 3: Monniken en ridders	40-70%	10-30%	20-50%
Historisch redeneren tijdvak 3			100%
Tijdvak 4: Steden en staten	40-70%	10-30%	20-50%
Historisch redeneren tijdvak 4			100%
...	...	...	...
Totaal			

Figuur 7.2. Toetsmatrijs geschiedenis

Er zijn tien blokken met vragen in de toets, 140 vragen in totaal. Per blok wordt een tijdvak getoetst. Vragen over historisch redeneren (verzamelen, ordenen, verklaren, beeld vormen) worden gesteld in de context van een tijdvak.

Alle items samen vormen een itembank waaruit vergelijkbare toetsen kunnen worden samengesteld. Bij elk item geven de constructeurs de verwachte moeilijkheid van de vragen aan en de tijd die de student nodig zal hebben om de vraag te beantwoorden. Deze informatie wordt door de hoofdredacteur gebruikt bij het samenstellen van de toets. De toets moet af te nemen zijn in twee uur.

7.5 De pretest en domeintoets: zicht op de kwaliteit van items

Om een eerste beeld te krijgen van de kwaliteit van de items, is in 2010 een pretest en in 2011 een domeintoets uitgevoerd. Voor de pretest zijn uit de itembank van ongeveer 500 items, handmatig 140 items geselecteerd op basis van de toetsmatrijs. Bij de domeintoets is een toets samengesteld uit alle items die betrekking hadden op een specifiek domein, te weten de eerste drie tijdvakken en items over historisch redeneren waarbij de nadruk ligt op het op de juiste manier informatie verzamelen. In 2012 zullen nog drie domeintoetsen en een pretest worden uitgevoerd.

Voor de afname van de toets wordt gebruik gemaakt van Question Mark Perception (QMP), een toetsservicesysteem waarbij de toetsen in een beveiligde omgeving direct op de computer afgenomen kunnen worden.

De pretest is in november 2010 bij 624 studenten (eerstejaars en derdejaars) van de lerarenopleidingen geschiedenis afgenomen en de domeintoets in november 2011 bij 46 studenten. De context voor deze eerste pretest was vrijblijvend: de pretest was niet verplicht voor studenten en het resultaat van de pretest had geen gevolgen voor de student. Daarnaast konden de studenten niet specifiek voor de pretest leren en waren de curricula van de opleidingen nog niet volledig gebaseerd op de kennisbasis. Als de studenten in 2013 aan de toets deelnemen moet dit natuurlijk wel gerealiseerd zijn.

De resultaten van de toetsen zijn na de afname door toetsdeskundigen geanalyseerd. Om een oordeel te geven over de kwaliteit van de pretest zijn de volgende vragen beantwoord:

1. Meet de toets wat hij beoogt te meten (validiteit)?
2. Meet de toets consistent (betrouwbaarheid)?
3. Hoe hebben de studenten op de toets gescoord?
4. Wat is de kwaliteit van de afzonderlijke items?

Bij de domeintoets is alleen gekeken naar de kwaliteit van de afzonderlijke items.

Op itemniveau is gekeken naar (a) de moeilijkheid van het item (p-waarde), (b) de afleiders en (c) het onderscheidend vermogen van het item. In Figuur 7.3 staat een voorbeeld van een itemanalyse waarbij het juiste alternatief een p-waarde heeft van 0,61 en een hoog onderscheid gemaakt wordt tussen de betere studenten (de bovenste groep) en de zwakkere studenten (de onderste groep).

Antwoord-mogelijkheden	Aantal keren gekozen	p-waarde	Bovenste groep	Onderste groep	Onderscheidend vermogen	Correlatie itemtoets
a*	28	0,61	0,92	0,17	0,75	0,53
b	4	0,09	0,08	0,00	0,08	0,08
c	4	0,09	0,12	0,08	0,04	-0,11
d	10	0,21	0	0,75	-0,75	-0,49
* juiste antwoord						

Figuur 7.3 Een itemanalyse voor geschiedenis

Items met een lage p-waarde zijn items die door veel studenten foutief zijn beantwoord. Deze items en de items met een laag of negatief onderscheidend vermogen tussen goede en zwakke studenten worden door de toetsdeskundige besproken met de hoofdredacteur. Op basis van deze analyse worden items bijgesteld en soms zelfs weggegooid.

7.6 De ervaringen van studenten

Na afloop van het maken van de toetsen evalueerden de studenten vervolgens het gemaakte werk. Bij de pretest in 2010 wisten studenten vooraf nog niet goed wat het doel van de toets was, bij de domeintoets is dit verbeterd. De studenten vinden de toets wel moeilijk en geven aan dat nog niet alle stof van de toets in het curriculum behandeld is. 95 Procent van de studenten heeft zich niet specifiek voor de domeintoets voorbereid. De studenten geven aan dat deze kennistoets helpt om de vakinhoudelijke kennis beter te beheersen. Ze vinden de toets wel lang en de teksten bij het historisch redeneren te uitgebreid. De toetsscore was voor eenderde van de studenten hoger dan verwacht en voor een even groot aantal studenten lager dan verwacht.

7.7 Conclusie

Het tussentijds aanscherpen van het ontwikkelproces en het uitvoeren van de pretest en domeintoets hebben een bijdrage geleverd aan het tot stand komen van kwalitatief goede items. De domeintoetsen die dit jaar

nog volgen zullen een compleet beeld geven van de relevantie, moeilijkheid en het onderscheidend vermogen van de items. Deze informatie zal er voor zorgen dat er goed onderbouwde keuzes gemaakt kunnen worden voor het vaststellen van een kwalitatief goede landelijke kennistoets.

In de context van *10voordeleraar* heeft het lectoraat de volgende producten opgeleverd, lezingen gegeven of anderszins een bijdrage geleverd

- Meta-analyse van het proces van kwaliteitsbewaking *10voordeleraar*.
- Rapportage van de domeintoetsanalyse Geschiedenis in de 3^e testcyclus.
- Handleiding interpreteren van toets- en item analyses.
- Lezing bij de startconferentie landelijke borging masteropleidingen: *Naar een kijkkader voor toetskwaliteit*, 12 oktober 2012.
- Workshops 'item- en toetsconstructie' bij de diversie vakken.
- Rapportages domeintoetsanalyses
- Vergelijking van de evaluatievragen over de vakken heen.
- Rapportages van de toets- en itemanalyse van de tweedegraads pretests
- Presentatie voor de Raad van Advies. Werkwijze van de kwaliteitscommissie.



Hoofdstuk 8

Trainen van beoordelingsvaardigheden

Suzanne Kratsborn

8.1 Samenvatting

Peer assessment is een toetsvorm die past bij de visie van Fontys PABO Tilburg (FPT), omdat zij past bij PGO, de ontwikkeling van meta-cognitieve vaardigheden bevordert, relevant is voor het beroep van leraar basisonderwijs en bijdraagt aan een goed toetsprogramma. Om peerassessments effectief en efficiënt in te zetten is het essentieel dat de studenten beschikken over beoordelingsvaardigheden. De training hiervoor moet geïntegreerd in het onderwijs worden opgenomen (Sluijsmans, 2002). Op FPT ontbreekt expertise op het gebied van het trainen van beoordelingsvaardigheden. In een literatuurstudie worden de volgende drie beoordelingsvaardigheden bestudeerd: het definiëren van criteria, beoordelen en het geven van feedback. Deze worden gerelateerd aan de volgende aspecten van de rol van de leraar: het belang van dialoog, het gebruik van voorbeelden en rubrics. Dit leidde tot kernconcepten rondom de beoordelingsvaardigheden en de rol van de leraar. In een curriculumonderzoek en interviews met tutoren per leerjaar wordt verkend of en in welke mate deze kernconcepten terug te vinden zijn in het curriculum van FPT en welke kansen er liggen in het huidige

curriculum. Zoals verwacht, werd er in het curriculum van FPT nog weinig aandacht besteed aan (aspecten van) beoordelingsvaardigheden. Wel werden er kansen gezien. Resultaat is een aanbeveling voor aanpassingen in het curriculum waardoor studenten getraind worden in beoordelingsvaardigheden ten behoeve van het effectief en efficiënt uitvoeren van peer assessment. Hierbij is aangesloten bij het huidige curriculum en gedefinieerde kansen.

8.2 Inleiding

Peer assessment is een toetsvorm waarbij studenten betrokken zijn bij het beoordelen van elkaar en waarbij studenten producten of andere opbrengsten van hun medestudenten evalueren, al dan niet gebruik makend van zelfgeselecteerde criteria (Dochy, Admiraal & Pilot, 2003; Sluijsmans, 2002; Topping, 2009).

Peer assessments bevorderen het hogere-orde-denken; de ontwikkeling van metacognitieve vaardigheden zoals reflecteren, feedback geven, analyseren, probleemoplossen en communicatie (Allal & Lopez, 2005; Sluijsmans, 2002). Deze houding en vaardigheden zijn

nodig om in het werkende leven te blijven leren (Van Zundert, 2012).

Peer assessment is zeer geschikt als ‘assessment for learning’: een integraal onderdeel van het leerproces in plaats van alleen een sluitstuk (Allal & Lopez, 2005, Topping, 2009; Wood, 2009). Door het assessment als een leermiddel te zien wordt de leeropbrengst verhoogd en diepgaand leren bevordert (Dochy et al., 2003; Sluijsmans, 2002). Peer assessment is een toetsvorm die kennisontwikkeling en leerprestaties bevordert, doordat zij effectieve feedback verschaft aan studenten (Dochy et al., 2003, Wood, 2009). Peerassessments geven ‘just in time’ en op het individu afgestemde feedback (Van den Berg, Pilot & Admiraal, 2005). Formatieve toetsing, waarbij het geven van ontwikkelingsgerichte feedback het doel is, verhoogt de motivatie en leidt tevens tot diepgaand leren (Dochy et al., 2003). Hoe sneller de feedback op de toets volgt, hoe directer de student zijn handelen kan bijstellen. Jaspers en Heijmen-Versteegen (2005) en Topping (2009) stellen dat wanneer studenten elkaar feedback geven, dit leerzaam is voor zowel de feedbackgever als de feedbackontvanger, waarbij peers meer elkaars taal spreken dan dat het geval is bij docenten en studenten. Het is belangrijk dat toetsing direct meegenomen wordt in het ontwerp van de leeromgeving.

Voor het beroep van leraar basisonderwijs is het inzetten van peer assessment waardevol omdat leraren basisonderwijs in hun werk ook in staat moeten zijn om anderen (kinderen) te beoordelen.

Om peer assessments effectief en efficiënt in te zetten is het essentieel dat de studenten beschikken over beoordelingsvaardigheden. In het pabocurriculum is voor de ontwikkeling van die beoordelingsvaardigheden nog weinig aandacht.

Beoordelingsvaardigheden worden gedefinieerd als vaardigheden die noodzakelijk zijn voor het beoordelen van een product of proces van jezelf of een ander (Sluijsmans, 2002). Van Zundert, Sluijsmans en Van Merriënboer (2010) en Van Zundert 2012), Sluijsmans (2002), Sluijsmans en Prins (2006) en Joosten-ten Brinke, Verhoeven en Van Buuren (2003) concluderen dat training in beoordelingsvaardigheden essentieel is voor de kwaliteit van peerassessments. Deze training moet geïntegreerd in het onderwijs opgenomen worden (Sluijsmans, 2002; Sluijsmans & Prins, 2006).

8.3 Onderzoeksvraag

Het volgende onderzoeksdoel is geformuleerd:
Een aanbeveling opleveren voor een training van beoordelingsvaardigheden in het curriculum van Fontys PABO Tilburg waardoor studenten in staat zijn peer assessments op een goede manier uit te voeren. Dit leidt tot de volgende onderzoeksvraag: Welke aanpassingen in het curriculum zijn nodig op Fontys PABO Tilburg om beoordelingsvaardigheden bij de studenten zo te ontwikkelen dat zij peer assessments effectief en efficiënt uit kunnen voeren?

Vanuit deze onderzoeksvraag zijn de volgende deelvragen opgesteld:

1. Welke vaardigheden liggen ten grondslag aan de beoordelingsvaardigheden die nodig zijn om een peer assessment goed uit te voeren?
2. Wat is de rol van de leraar bij het aanleren van de beoordelingsvaardigheden?
3. Wat wordt er op Fontys PABO Tilburg al gedaan op het gebied van beoordelingsvaardigheden?
4. Welke kansen liggen er in het huidige curriculum om training in beoordelingsvaardigheden in te zetten?

8.4 Methoden

De onderzoeksmethoden die zijn gebruikt zijn een literatuuronderzoek (deelvraag 1 en 2), curriculumonderzoek en interviews (deelvraag 3 en 4).

Literatuurstudie

Door een literatuurstudie worden kernconcepten gedefinieerd op het gebied van beoordelingsvaardigheden (deelvraag 1) en de rol van de leraar hierbij (deelvraag 2). De gehanteerde zoektermen zijn: peer assessment, beoordelingsvaardigheid, criteria definiëren, beoordelen, feedback geven. Er is gezocht binnen de disciplines psychologie en onderwijswetenschappen/ sociale wetenschappen naar artikelen in het taalgebied Engels en Nederlands van 1990-heden. Er is gebruik gemaakt van de digitale bibliotheken van Fontys en ESoE. Er is gezocht in titel en samenvatting. Tevens is gebruik gemaakt van de sneeuwbalmethode, waarbij de referentielijst van een

gevonden artikel is gescreend op bruikbare bronnen. In totaal zijn er 45 artikelen geselecteerd.

Het resultaat van de literatuurstudie is een beargumenteerd voorstel voor de te ontwikkelen beoordelingsvaardigheden en richtlijnen voor de manier waarop deze vaardigheden ontwikkeld kunnen worden.

Curriculumonderzoek

Curriculumonderzoek geeft zicht op onderdelen van het curriculum waarin aandacht is voor de gedefinieerde kernconcepten.

De volgende documenten per leerjaar worden gescreend op de geformuleerde kernconcepten: overzichten van leerarrangementen en bijhorende tutorhandleidingen, toetsoverzichten en beoordelingsprotocollen.

Naar aanleiding van het bestudeerde overzicht van leerarrangementen zijn de powerpointpresentaties van drie hoorcolleges gescreend, omdat in het overzicht een kernconcept werd gerelateerd aan dit hoorcollege. De documenten geven samen een goed beeld van de inhoud van het curriculum. In totaal zijn 32 documenten gescreend. Letterlijke tekstfragmenten zijn per kernconcept genoteerd.

Het curriculum is vanuit twee oogpunten bekeken: ten eerste wordt beschreven welke kernconcepten in het curriculum al aanwezig zijn (deelvraag 3) en ten tweede op welke momenten aandacht voor beoordelingsvaardigheden uitgebreid of logischerwijs ingevoerd kan worden (deelvraag 4).

Interviews

Interviews zijn om twee redenen gehouden.

1. Er gebeurt in het onderwijs meer dan op papier te vinden is. In de documenten is niet altijd te vinden of onderwerpen op dusdanige wijze aandacht krijgen als in de kernconcepten van dit onderzoek gevraagd is. De interviews vergroten de data.
2. In het kader van triangulatie zijn de gegevens uit de documentanalyse gecheckt.

Het interview is een gedeeltelijke gestructureerd interview: De onderwerpen liggen vast, dat zijn de geformuleerde kernconcepten op basis waarvan ook de documentenanalyse heeft plaats gevonden. Vragen zijn geformuleerd per kernconcept. Dat zijn zowel vragen waarin de juistheid van de documentenanalyse gecheckt wordt en als verhelderende vragen waarin doorgevraagd wordt op operationalisering van in documenten gevonden beschrijvingen. Op basis van gegeven antwoorden wordt doorgevraagd. De drie beoordelingsvaardigheden met de bijhorende kernconcepten bepalen de volgorde van het interview. Indien nodig kan de volgorde van de te bespreken kernconcepten worden aangepast aan het verloop van het interview. Steeds wordt gestart met een open vraag over de beoordelingsvaardigheid. Hierdoor kan de geïnterviewde alles antwoorden wat volgens hem/haar van toepassing is. Daarna wordt doorgevraagd op de onderliggende kernconcepten.

Interviews vinden plaats met twee (van de vier tot zes) tutoren per opleidingsjaar, in elk geval is één van deze

personen ook coördinator van een leerarrangement. Beide personen hebben een goede kijk op het onderwijs in het betreffende opleidingsjaar. Ze zijn er meerdere jaren tutor. De keuze om de twee personen tegelijk te interviewen is gemaakt omdat het gaat om feitelijkheden in een curriculum en geen gevoelige informatie betreft. Door de twee personen tegelijk te interviewen kunnen zij op elkaar reageren waardoor de verwachting is dat zij door inbreng van de ander op meer ideeën komen. Op deze manier worden meer data gegenereerd.

Er wordt gebruik gemaakt van registratieapparatuur en de interviewer maakt korte notities tijdens de interviews. Achteraf worden de interviews per kernconcept uitgeschreven. De letterlijke tekstfragmenten met betrekking tot de kernconcepten wordt naast de documentanalyse gezet, zodat interpretatie van gegevens zo lang mogelijk wordt uitgesteld. In dit hoofdstuk zullen de resultaten kort worden weergegeven.

8.5 Resultaten

Belangrijkste bevindingen uit het literatuuronderzoek naar beoordelingsvaardigheden en de rol van de leraar: Beoordelingsvaardigheden zijn vaardigheden die noodzakelijk zijn voor het beoordelen van een product of proces van jezelf of een ander (Sluijsmans, 2002). Van Zundert (2010) en Sluijsmans (2002) concluderen dat een training in beoordelingsvaardigheden essentieel is voor de kwaliteit van peer assessment. Als studenten deze beoordelingsvaardigheden beheersen zijn zij in

staat om peerassessments met voldoende zelfvertrouwen uit te voeren, de kwaliteit van de gegeven feedback te verhogen en ook inhoudelijk leerrendement te verkrijgen. Sluijsmans (2002) geeft aan dat de volgende drie beoordelingsvaardigheden relevant zijn voor peer assessment:

1. *Het definiëren van assessmentcriteria*

Vooraf bedenken wat relevante elementen zijn bij beoordeling van het product en/of proces. Aanbevolen wordt om studenten te betrekken bij het definiëren van criteria, omdat dit studenten meer inzicht in het te maken product geeft.

2. *Het beoordelen*

De student moet hiervoor het werk van een andere student kunnen analyseren op basis van de beoordelingscriteria.

3. *Het geven van feedback*

De student moet feedback op een constructieve wijze kunnen geven en zowel sterke punten als ontwikkelpunten benoemen in de feedback.

Uit literatuurstudie (o.a. Cartney 2010; Goodrich, 2005; Nicol & Macfarlane-Dick, 2006) komt naar voren dat bij het aanleren van beoordelingsvaardigheden drie aspecten in de rol van de leraar cruciaal zijn, te weten het belang van dialoog, het werken met voorbeelden en het werken met rubrics.

1. *Het belang van dialoog*

Studenten ervaren dat een open toetsdialoog gecreëerd wordt. Het delen van gemeenschappelijke ervaringen en het bediscussieren ervan maakt het mogelijk om meer kennis expliciet te delen. Vanuit deze kennisdeling ontstaat een gemeenschappelijke context en daarmee een gemeenschappelijk beeld van gedrag in een context dat nodig is om te voldoen aan de doelen en criteria (Cartney, 2010).

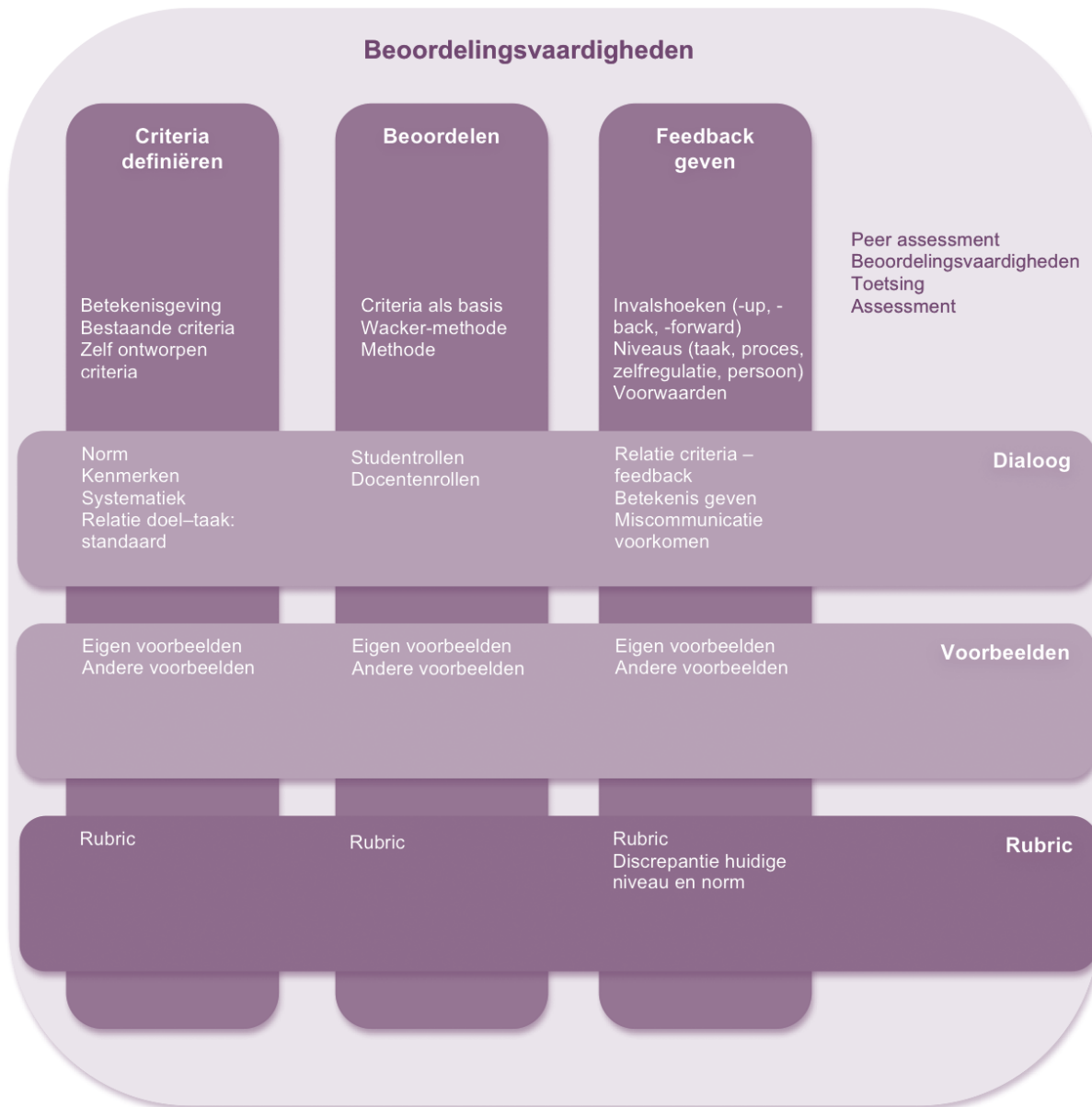
2. *Het werken met voorbeelden*

Voorbeelden vormen de spil in het aanleren van beoordelingsvaardigheden: zij zijn het middel om de criteria in gemaakt werk te herkennen en er vervolgens feedback op te geven (Nicol & Macfarlane-Dick, 2006).

3. *Het werken met rubrics*

Een manier om gedrag in een bepaalde context vast te leggen is het gebruik van een rubric. Hierin wordt observeerbaar gedrag gekoppeld aan de criteria (Goodrich, 2005).

Vervolgens zijn de beoordelingsvaardigheden in relatie tot de rol van de leraar gebracht. Dat leverde een overzicht van kernconcepten op. Deze kernconcepten vormden de basis voor het curriculumonderzoek en de interviews.



Figuur 8.1 Beoordelingsvaardigheden in Relatie tot de Rol van de Leraar. Gedefinieerde Kernconcepten.

In een curriculumonderzoek en interviews met tutoren per leerjaar werd verkend of en in welke mate deze kernconcepten terug te vinden zijn in het curriculum van FPT en welke kansen er liggen in het huidige curriculum.

Er vinden op verschillende momenten in het curriculum van Pabo Tilburg activiteiten plaats die passen binnen het trainen van beoordelingsvaardigheden, maar dat gebeurt niet vanuit het perspectief van het leren van beoordelingsvaardigheden. Het gebeurt veelal ad hoc en daar waar peer assessment ingezet wordt is er geen sprake van het aanleren van de afzonderlijke vaardigheden. Er liggen kansen om deze momenten in het geheel van een training in beoordelingsvaardigheden te bezien en uit te bouwen tot structurele onderdelen van het curriculum.

8.6 Conclusie

Resultaat is een aanbeveling voor aanpassingen in het curriculum waardoor studenten getraind worden in beoordelingsvaardigheden ten behoeve van het effectief en efficiënt uitvoeren van peer assessment. Hierbij is aangesloten bij het huidige curriculum en gedefinieerde kansen.

In het onderzoek zijn per leerjaar concrete voorstellen gedaan voor het implementeren van de beoordelingsvaardigheden. Per beoordelingsvaardigheid komen de aspecten van de rol van de leraar (dialoog, voorbeelden en rubrics) terug.

Peer assessment

Zowel uitvoerende als schriftelijke toetsmomenten als stagemomenten zijn geschikt voor het uitvoeren van peer assessment, omdat studenten dan gericht vanuit criteria aan het werk zijn. Binnen de kernproducten wordt regelmatig groepswork verondersteld. Ook het samenwerkingsproces kan onderdeel van training in beoordelingsvaardigheden zijn. Op dit moment wordt er al feedback gegeven op de samenwerking, echter niet systematisch en vanuit criteria.

In de propedeuse lopen studenten (waar mogelijk) in tweetallen stage in een klas. Zij werken toe naar een Performance Assessment (PA1) met vastgestelde criteria. In hoofdfase2 lopen studenten soms stage in tweetallen en (waar mogelijk) met meerdere studenten op een stageschool. In hoofdfase2 vinden PA2 en PA3 plaats volgens vastgestelde criteria. Situaties waarin praktische voorwaarden voor feedback geven aan elkaar er zijn.

In de literatuur over beoordelingsvaardigheden staat nergens dat de volgorde waarin aandacht besteed wordt aan de drie beoordelingsvaardigheden vast ligt. Wel blijkt dat er verband is tussen de vaardigheden. Bijvoorbeeld: door het beoordelen wordt begrip van de criteria groter. Begrip van criteria is een vorm van feed-up (vorm van feedback) en vormt de context waarin feedback gezien moet worden (Handley & Williams, 2011).

Criteria definiëren

In hoofdfase2 en de afstudeerfase was er in het vorige studiejaar op enkele momenten aandacht voor het zelf formuleren van criteria. Uit het literatuuronderzoek blijkt overduidelijk dat het belangrijk is om studenten te betrekken bij het formuleren van criteria (zie onder andere Castelijns, Segers, & Struyven, 2011; Sluijsmans, 2002). Ook wordt gesteld dat eerstejaars studenten nog weinig voorkennis en gemeenschappelijke ervaringen hebben waardoor er weinig expliciete kennis is. Hierbij is een inductief-presenterende manier (Janssen-Noordman en Van Merriënboer, 2009) van werken wenselijk. Het begrip van de taak wordt groter door vanuit voorbeelden te werken. Dit is een noodzakelijke voorwaarde om criteria te kunnen definiëren (Handley & Williams, 2011).

Door een rubric te gebruiken worden voorbeelden gegeven van allereerst het gebruik van een rubric 'an sich', maar ook van de norm: het gedrag dat de student moet laten zien om te voldoen aan een criterium. In een opbouw zou dat er als volgt uit kunnen zien: als eerste worden criteria en rubric gegeven, vervolgens worden alleen criteria gegeven en wordt de rubric door studenten gemaakt en ten slotte worden zowel criteria als rubric door studenten gemaakt. Vanaf het eerste moment moet er aandacht zijn voor de betekenis van de criteria waarbij een koppeling aan voorbeelden, allereerst van andere studenten, waarin de criteria duidelijk herkenbaar zijn van belang is. De rubric wordt op deze manier ingezet als leermiddel (Goodrich, 2005).

De rol van de docent zal van een inductief-presenterende manier naar een inductief-vragende manier moeten veranderen: bij beiden vormt het werken vanuit voorbeelden het uitgangspunt. De rol van de docent verandert van het presenteren naar het studenten vragenderwijs gemeenschappen laten ontdekken.

Beoordelen

Nicol en Macfarlane-Dick (2006) geven aan dat studenten door begrip van de criteria en doelen en het oefenen met feedback geven een beeld ontwikkelen van het verschil in prestatie van het werk van een student en de gevraagde prestatie. Castelijns et al. (2011) en Sluijsmans (2002) geven aan dat de eerste stap in het beoordelen het vaststellen van het bewijs voor aanwezigheid van de criteria in het werk van de student is. Het beschrijven van de discrepantie tussen het feitelijke en gewenste niveau is een volgende stap.

Methodisch werken is noodzakelijk om objectievere manier van werken te borgen (QRM, 2010). De WACKER-methode biedt de kans om systematisch te beoordelen: Waarnemen (W) en aantekeningen maken (A) vormen de basis. De aantekeningen classificeren (C) en kwalificeren (K) is bij duidelijke criteria en rubric het meest eenvoudig (indien aan de voorwaarde 'begrip van de criteria' is voldaan). In eerste instantie zal het ook zeker bijdragen aan het begrijpen van de criteria.

De dialoog tussen studenten speelt een belangrijke rol in het evalueren (E). Het helpt de analyse als studenten hun oordeel moeten toelichten aan andere studenten.

Studenten en de docent stellen vragen over het oordeel (Nicol & Macfarlane-Dick, 2006).

De rapportage (R) is onderdeel van het geven van feedback.

Het werken met voorbeelden in deze fase van het peer assessment is van waarde voor het herkennen van criteria in gemaakt werk en er vervolgens feedback op te geven (Nicole & Macfarlane-Dick, 2006).

Feedback geven

In de propedeuse is er aandacht voor het feedback geven aan elkaar. Door dit uit te bouwen en te koppelen aan peer assessment wordt een bestaand element versterkt. In de volgende jaren zou dit uitgediept moeten worden. Uit de literatuurstudie blijkt dat feedback gericht op de taak de basis vormt voor verdere feedback (Hattie & Timperly, 2007). Hier zou men zich in de propedeuse op moeten richten. In hoofdfase2 kan een uitbreiding gemaakt worden op feedback op het proces. Het derde niveau is feedback op zelfregulatie. Een hogere mate van zelfregulatie wordt al bereikt door te werken met peer assessment (Nicol & MacFarlane, 2006).

In hoofdfase3 en de afstudeerfase kan hierop voortgebouwd worden.

Miller (2003) geeft aan dat feedback moet verwijzen naar een specifiek punt in het werk van de medestudent. In de WACKER-methode wordt bij het 'aantekeningen maken' beschreven dat de studenten punten in het werk van de

medestudent moeten markeren die bewijs vormen voor de aanwezigheid van criteria. Dit is een heel concrete manier om feedback te geven. Vervolgens moet dit onderwerp van gesprek zijn in de groep om miscommunicatie te voorkomen (Cartney, 2010).

Zowel voorbeelden van andere (eerdere groepen) studenten als van de studenten zelf dienen gebruikt te worden. Zij dienen ieder een ander doel.

Gebruik van een rubric is hier van meerwaarde, omdat het een niveau koppelt aan de feedback. In de propedeuse geeft het handvatten om heel concreet feedback te geven: Punten in het werk van de student worden gemarkeerd en gekoppeld aan een criterium en niveau-aanduiding in de rubric. Of de rubric wordt samengesteld door studenten of docenten heeft te maken met het definiëren van criteria.

Per leerjaar is een opbouw gemaakt in de aan te leren beoordelingsvaardigheden. Per beoordelingsvaardigheid komen de aspecten van de rol van de leraar (dialog, voorbeelden en rubric) terug. De inhouden zijn cumulatief: aspecten uit voorgaande leerjaren worden bekend verondersteld in volgende leerjaren.

Het trainen van beoordelingsvaardigheden is geconcentreerd in de propedeuse en hoofdfase2. In hoofdfase3 en de afstudeerfase worden eerder verworven vaardigheden geconsolideerd.

In de propedeuse wordt gewerkt vanuit door de opleiding vastgestelde criteria. Er vindt training in het beoordelen plaats. Deze wordt vanaf hoofdfase2 herhaald. In de propedeuse wordt gestart met het trainen van het geven van feedback. Dit wordt in hoofdfase2 uitgediept.

In hoofdfase2 wordt gestart met het definiëren van criteria (en maken van een rubric). De manier waarop de docent in dialoog gaat met de student verandert in hoofdfase3 van inductief-presenterend naar inductief-vragend.

Het onderzoek is een beschrijvend onderzoek. Uitbreiding met een experimenteel onderzoek zou de betrouwbaarheid vergroten bij het bepalen of het uitgewerkte voorstel voor een training effectief is.

Peer assessment hangt nauw samen met self-assessment. In dit onderzoek is self-assessment buiten beschouwing gelaten.

8.7 Discussie

De documentenanalyse heeft voldoende data en basis voor de interviews opgeleverd. Als de documentenanalyse ook alle handleidingen (lesvoorbereidingen) van alle vakdocenten voor hun vakuren had omvat was wellicht het aantal data vergroot. De acht tutoren die geïnterviewd zijn hebben ook ieder hun eigen vakgebied, maar zijn in het interview bevraagd vanuit hun tutorrol.

Dit onderzoek is uitgevoerd op één Pabo. Indien het onderzoek op meerdere Pabo's zou zijn uitgevoerd, zouden meer data gegenereerd worden. Op basis van deze data zou een voorstel voor training van beoordelingsvaardigheden op Pabo's geformuleerd kunnen worden. Het zou ook meer zicht geven op aandacht voor beoordelingsvaardigheden op Nederlandse Pabo's. Eventueel zou zelfs het hele gebied van lerarenopleidingen (ook voor het V.O.) gebruikt kunnen worden voor een (grootschalig) onderzoek.

Hoofdstuk 9

Bewust en bekwaam toetsen en beoordelen

9.1 Samenvatting

Het SURF-project 'Bewust en bekwaam toetsen' heeft gedurende de gehele eerste periode van het lectoraat gelopen. Dit hoofdstuk is geschreven op basis van het onderzoeksrapport van Desirée Joosten-ten Brinke, Dominique Sluijsmans, Sabine van Eldik, Anneriet Florack, Herman Schalk en Tamara van Schilt-Mol. Bewust en bekwaam toetsen. De effecten van ontwikkelingsgerichte zelfbeoordeling door lerarenopleiders op hun toetskennis en toetskunde. Dit rapport is geschreven in het kader van het project Bewust en bekwaam toetsen en beoordelen dat uitgevoerd is in het kader van het SURF-programma Toetsing en Toetsgestuurd Leren. Het onderzoeksproject had de volgende drie doelstellingen: (1) Het ontwerp en ontwikkeling van een zelfbeoordelingsinstrument rondom eigentijds (digitaal) toetsen en beoordelen; (2) Lerarenopleiders bewust maken van de eigen toetskennis en -kunde; en (3) Verhoogde kwaliteit van toetsontwerpen en meer zicht op professionaliseringsdoelen van lerarenopleiders. Het doel van het onderzoek is na te gaan of het gebruik van een zelfbeoordelingsinstrument waarmee lerarenopleiders

plaats- en tijdonafhankelijk hun kennis in kaart kunnen brengen en professionaliseringsbronnen ontvangen leidt tot een verhoging van hun toetskennis.

In het project participeren vier instellingen:

- Fontys Lerarenopleiding Tilburg, Fontys Hogeschool;
- Het Instituut voor Leraar en School (ILS) van de Faculteit Educatie, Hogeschool van Arnhem en Nijmegen
- De Nieuwste PABO, Hogeschool Zuyd;
- De Universitaire Lerarenopleiding (ULO) van de Vrije Universiteit Amsterdam.

Op basis van analyse van bestaande instrumenten is een ontwikkelingsgericht zelfbeoordelingsinstrument ontwikkeld in samenwerking met leraren en assessmentexperts. Het instrument is te vinden op www.bewustenebekwaamtoetsen.nl.

Aan het onderzoek in dit project hebben 99 lerarenopleiders van vier lerarenopleidingen deelgenomen: ULO/VU Amsterdam, ILS/Hogeschool van Arnhem en Nijmegen, Fontys Lerarenopleiding

Tilburg/Fontys Hogescholen en De Nieuwste PABO/Zuyd Hogeschool. Het onderzoek naar het gebruik van het zelfbeoordelingsinstrument heeft plaatsgevonden in de periode van 14 januari tot en met 26 februari 2013.

De lerarenopleiders namen voor het onderzoek deel aan een scholings- en onderzoeksbijeenkomst waar ze een kennistoets over toetsen en beoordelen maakten en waarbij ze een casustoets maakten die gerelateerd was aan een van de lagen van de kwaliteitspiramide voor eigentijds toetsen en beoordelen (zie figuur 1.1). Door deze casustoets eerst te maken zonder gebruik te maken van het zelfbeoordelingsinstrument en daarna met gebruik van het instrument is het effect van dit instrument op de kwaliteit van de uitwerking onderzocht. Tevens is een evaluatievragenlijst ingevuld om te onderzoeken in welke mate de lerarenopleiders het instrument waardeerden.

Het onderzoek laat zien dat het gebruik van het instrument bijdraagt aan de prestatie op de casustoets. De evaluatie van het instrument toonde aan dat het instrument zinvol, authentiek en betekenisvol is en het de zelfbeoordeling stimuleert. Docenten gaven aan dat het instrument hen bewust had gemaakt van de omvang van het domein van toetsen en beoordelen. Zonder specifieke scholingscontext wordt het instrument echter niet meer gebruikt door de lerarenopleiders. Het onderzoek heeft zich vooral gericht op de toetskennis van lerarenopleiders.

9.2 Aanleiding en probleemstelling

Kwaliteit van toetsen en beoordelen is in de hele onderwijskolom een belangrijk thema. Het leveren van kwaliteit is echter moeilijk en voor een belangrijk deel afhankelijk van de kwaliteit van de personen die betrokken zijn bij toetsen en beoordelen. Toetsing is een ontwerpvragestuk, waarbij de aansluiting met het leren essentieel is (Biggs, 1996; Schuwirth & Van der Vleuten, 2006). Dit betekent dat toetsing naast de gangbare summatieve (=selecterende, beslissende) functie ook een formatieve (=ontwikkelingsgerichte) functie heeft. Bij deze ontwikkelingsgerichte functie van toetsing staat de bijdrage die de toetsing levert aan het leerproces van de lerende centraal (Sluijsmans, Joosten-ten Brinke, & Van de Vleuten, 2013). Als gesproken wordt over kwaliteit van toetsen gaat het niet uitsluitend om de kwaliteit van het meetinstrument. Toetsing vormt de ruggengraat in het curriculum van de opleiding en is daarmee een sterk sturende factor in het bepalen van kwaliteit van het onderwijs. Studenten oordelen nog steeds vaak negatief over de wijze van toetsing, toetsing kost in het algemeen veel tijd (10% van totale ontwerp- en ontwikkeltijd) en nog veel te vaak blijft het ontwerp van goede toetsen het ondergeschoven kindje van het onderwijs. In het hoger onderwijs is geconstateerd dat het met de kennis en kunde van (leraren)opleiders over toetsen niet goed is gesteld (Rixsen, 2010). Deze mening is niet ongegrond: lerarenopleiders zijn niet of nauwelijks opgeleid in goede toetsing en hebben onvoldoende kennis van de basisprincipes van toetsen en beoordelen (ISO, 2009;

Straetmans, 2006). Lerarenopleiders hebben onvoldoende zicht op hun kennis en kunde over toetsing. Dit is zorgelijk, omdat het bij toetsing gaat om het vaststellen van een betrouwbaar beeld van de kennis en kunde van de lerende en er vaak ingrijpende beslissingen aan worden verbonden.

Lerarenopleiders leiden studenten op tot leraar primair onderwijs, voortgezet onderwijs of middelbaar beroepsonderwijs die, net als de lerarenopleider zelf, voldoende toetsbekwaam moeten zijn om kwaliteit van toetsing te realiseren. Het opleiden van leraren gebeurt vanuit het congruentieprincipe waarbij de didactiek die door de lerarenopleiding wordt gestimuleerd, voorbeeldig is voor de didactiek die gebruikt wordt in de reële werksituatie (Koster & Korthagen, 1997; Swennen, Lunenberg, & Korthagen, 2008). Lerarenopleiders moeten het goede voorbeeld geven aan de studenten wat betreft de inrichting en vormgeving van het eigen toetsbeleid, het beoordelen van de studenten en het reflecteren op eigen handelen. Studenten leren hierdoor impliciet hoe kwalitatief goed toetsen en beoordelen eruit ziet. Om ervoor te zorgen dat studenten dit voorbeeldgedrag ook herkennen als goed voorbeeld is het nodig dat de keuzes geëxpliciteerd worden en gelinkt worden aan relevante theorieën (Swennen e.a., 2008). Een goede lerarenopleider leert studenten expliciet hoe je als leraar en als school(team) goed toetst in een specifieke onderwijscontext (Joosten-ten Brinke, 2011). Om deze voorbeeldfunctie goed uit te voeren vereist dit van de lerarenopleiders dat zij beschikken over een goede kennisbasis over toetsing.

Het besef dat de kwaliteit van toetsing voor een groot deel de kwaliteit van het onderwijs vertegenwoordigt en dat daarom leren en toetsen goed op elkaar aan moet sluiten, heeft ertoe geleid dat de professionaliseringsactiviteiten rondom toetsing de laatste jaren als paddenstoelen uit de grond schieten. Er zijn studiedagen en conferenties over toetsing, er zijn expertgroepen die opleidingen adviseren en ook vanuit onderzoek vindt kennisuitwisseling over toetsing plaats. Toch blijkt de huidige professionalisering te weinig effectief en efficiënt. Een belangrijke oorzaak hiervoor is dat professionaliseringsbijeenkomsten veelal worden gepland vanuit een aanbodgericht perspectief. Professionalisering 'on demand' is lastig en nog niet uitgewerkt, waardoor er voor just-in-time vragen te weinig ruimte is. Professionaliseringsbijeenkomsten sluiten daardoor niet aan bij het bestaande kennisniveau van de deelnemers. Een bijeenkomst is voor de ene opleider teveel herhaling, terwijl voor een andere opleider de stappen te groot zijn. Daarnaast is professionalisering veelal te instrumenteel en gericht op de kwaliteit van een specifieke toetsen. Er kan echter pas sprake zijn van kwaliteit van toetsing als deze vanuit een breed perspectief benaderd wordt, waarbij ook een toetsbeleid waarin het duurzaam effect van toetsen gepositioneerd is en de kwaliteit van toetsprogramma's wordt meegenomen (Joosten-ten Brinke, 2011). Mogelijke oplossingen om de professionalisering effectiever en efficiënter te maken zouden kunnen liggen in het meer vraaggestuurd en ontwikkelingsgericht aanbieden van de professionalisering. Denis en Van Damme (2009) en

Desimone (2009) noemen de volgende eisen waar een instrument met een professionaliseringsdoel aan moet voldoen:

- het instrument stimuleert een onderzoekende houding (actie-onderzoek);
- het instrument heeft een focus op het leren van professionals;
- het instrument levert feedback op ten aanzien van het eigen kennis- en vaardigheidsniveau in relatie tot wenselijke resultaten;
- het instrument levert informatie over de richting waarin de eigen kennis en vaardigheden kunnen worden ontwikkeld;
- het instrument kan individueel 'just-in-time' en in een groepscontext ingezet worden;
- vraagvormen worden (deels) digitaal aangeboden;
- vragen zijn gericht op eigen ervaringen: welke activiteiten heb je ondernomen, op welke manieren heb je kennis vergaard, etc.;
- er wordt gebruik gemaakt van bestaande toetsvormen;
- de aanpak is vraaggestuurd, waarbij het stimuleren van eigenaarschap centraal staat.

Opleidingen beschikken echter niet over zo'n ontwikkelingsgericht instrument waarmee lerarenopleiders plaats- en tijdonafhankelijk hun kennis en kunde over toetsing kunnen meten en op basis daarvan zich verder kunnen ontwikkelen. Het doel van dit onderzoek is te onderzoeken of het gebruik van een instrument waarmee lerarenopleiders plaats- en tijdonafhankelijk hun kennis

in kaart kunnen brengen en professionaliseringsbronnen ontvangen leidt tot een verhoging van hun toetskennis.

Aangezien het gaat om plaats- en tijdonafhankelijk ligt het voor de hand om te kiezen voor een digitaal instrument waarmee de opleiders zichzelf kunnen beoordelen.

De volgende vragen staan in het onderzoek centraal:

1. Wat is het effect van een digitaal zelfbeoordelingsinstrument voor toetsen en beoordelen op de toetskennis en -kunde van lerarenopleiders?
2. Hoe evalueren lerarenopleiders een digitaal zelfbeoordelingsinstrument voor toetsen en beoordelen?

Om deze onderzoeksvragen te kunnen beantwoorden is het nodig om te beschikken over een zelfbeoordelingsinstrument voor toetsen en beoordelen. In paragraaf 9.3 staat het ontwerp en de ontwikkeling van dit instrument. Daarna volgen in paragraaf 9.4 de resultaten en conclusies van het onderzoek.

9.3 Het ontwerp en de ontwikkeling van het zelfbeoordelingsinstrument

Om vast te kunnen stellen welke toetskennis en -kunde van lerarenopleiders moet worden verwacht is een kader nodig om deze kennis en kunde te positioneren. De kwaliteitspiramide van eigentijds toetsen en beoordelen

heeft als basis gediend voor het ontwerp van het zelfbeoordelingsinstrument.

Bij de start van dit onderzoeksproject was er onvoldoende zicht op wat een lerarenopleider zou moeten weten over toetsing. Beschrijvingen over de gewenste kennis en kunde zijn onvoldoende vertaald in leerdoelen (zie bijvoorbeeld Bekker-Holtland en anderen, 2010; Tillema, Dirkse-Hulsher, & Van den Hurk, 2002; Van Dinther, 2009). Ook in de beroepsstandaard voor lerarenopleiders is de bekwaamheidseis voor toetsing te abstract omschreven (Koster & Dengerink, 2000): 'De lerarenopleider ontwikkelt instrumenten voor (zelf)beoordeling van professionele bekwaamheden en beoordeelt deelnemers op hun geschiktheid'. Er is echter niet uitgewerkt welke

kennis hieraan ten grondslag ligt. In de kennisbasis voor lerarenopleiders (<http://www.velon.nl/kennisbasis>) is met name binnen de domeinen 'leren en lerenden' en 'curriculum en assessment' aandacht voor toetsen en beoordelen. Het draait hier om de vragen 'wanneer heeft iemand iets geleerd' en 'hoe toets je of de doelen van het curriculum worden bereikt'. Ook in een reviewstudie over de rollen van leraren staat de rol van beoordelaar niet expliciet benoemd (Lunenberg, Dengerink, & Korthagen, 2013).

Op basis van een literatuuronderzoek en interviews met assessmentexperts is in kaart gebracht wat de kennis en kunde van lerarenopleiders zou moeten zijn (zie Tabel 9.1; Sluijsmans, Van Eldik, Joosten-ten Brinke, & Jakobs, 2013).

Tabel 9.1. Een overzicht van de benodigde toetskennis en toetskunde van lerarenopleiders.

Leerdoelen op het niveau van toetsvragen
1. De lerarenopleider kent verschillende typen toetsvragen. 2. De lerarenopleider weet op welke beheersingsniveaus een toetsvraag zich richt. 3. De lerarenopleider kan op basis van leerdoelen toetsvragen construeren die voldoen aan de kwaliteitscriteria relevantie, objectiviteit, efficiëntie en juiste moeilijkheidsniveau (experts). 4. De lerarenopleider kan toetsvragen op de kwaliteitscriteria relevantie, objectiviteit, efficiëntie en juiste moeilijkheidsniveau beoordelen. 5. De lerarenopleider kan realistische, concrete situaties en authentieke casussen bedenken. 6. De lerarenopleider kan op basis van beoordelingscriteria gegeven antwoorden beoordelen. 7. De lerarenopleider kan studenten scoren op individuele toetsvragen. 8. De lerarenopleider kan toetsvragen na afname analyseren op psychometrische kwaliteit.

Vervolg tabel 9.1. Een overzicht van de benodigde toetskennis en toetskunde van lerarenopleiders.

Leerdoelen op het niveau van toetsen

1. De lerarenopleider kan een toets ontwerpen.
2. De lerarenopleider weet wat een toetsmatrijs is en hoe deze opgesteld moet worden.
3. De lerarenopleider kent de kwaliteitscriteria van toetsinstrumenten en weet deze toe te passen en te analyseren.
4. De lerarenopleider kan toetsbare leerdoelen (gebaseerd op inhoud en beheersingsniveau) voor studenten formuleren.
5. De lerarenopleider weet hoe fraude en plagiaat voorkomen kunnen worden.
6. De lerarenopleider kent verschillende toetsvormen en kan een toetsvorm kiezen en verantwoorden die aansluit bij de doelstellingen.
7. De lerarenopleider kent de voor- en nadelen van verschillende toetsvormen.
8. De lerarenopleider kent het verschil tussen proces- en productbeoordeling (experts).
9. De lerarenopleider kent de functies van toetsing (experts).
10. De lerarenopleider weet wat pre- true- en post-effecten van toetsen zijn.
11. De lerarenopleider kan studenten betrekken in het toetsproces en hoe zij zichzelf en elkaar kunnen beoordelen.
12. De lerarenopleider kan een selecterende toets construeren die de 'goede' studenten onderscheidt van de 'zwakkere' studenten.
13. De lerarenopleider kan toetsen aanpassen aan het niveau van de studenten.
14. De lerarenopleider kan heldere beoordelingscriteria construeren, en een rubric ontwikkelen.
15. De lerarenopleider kent de beoordelaarseffecten (zoals het halo-effect, het volgorde-effect).
16. De lerarenopleider kan de juiste toetsomgeving voor afname inrichten.
17. Een lerarenopleider kent het begrip cesuur en kent methoden voor het vaststellen van de cesuur en kan de cesuur bepalen.
18. De lerarenopleider kan een scoremodel opstellen en cijfers toekennen.
19. De lerarenopleider kan toetsafnames organiseren en uitvoeren.
20. De lerarenopleider weet wat intersubjectief beoordelen is.
21. De lerarenopleider kan gedrag in authentieke situaties beoordelen.
22. De lerarenopleider kan een antwoordmodel construeren (experts).
23. De lerarenopleider kent het verschil tussen analytisch en holistisch beoordelen.
24. De lerarenopleider kan na afname van een toets de kwaliteit van de toets vaststellen door een kwantitatieve en kwalitatieve toetsanalyse (experts).
25. De lerarenopleider kan toetsresultaten gebruiken voor verbetering van het onderwijs.
26. De lerarenopleider kan een gegeven beoordeling beargumenteren en rechtvaardigen en uitleggen.
27. De lerarenopleider kan effectieve, constructieve en bruikbare feedback geven (experts).
28. De lerarenopleider kan kwantitatieve toetsresultaten omzetten naar betekenisvolle informatie (experts)

Vervolg tabel 9.1. Een overzicht van de benodigde toetskennis en toetskunde van lerarenopleiders.

Leerdoelen op het niveau van toetsprogramma's

1. De lerarenopleider weet wat een toetsprogramma is.
2. De lerarenopleider weet hoe toetsen in een toetsprogramma samenhangen in relatie tot de leerdoelen.
3. De lerarenopleider kan verantwoorden hoe de resultaten op afzonderlijke toetsen bijdragen aan de eindkwalificaties van de opleiding.
4. De lerarenopleider kent kwaliteitscriteria voor toetsprogramma's en kan ze toepassen.
5. De lerarenopleider weet wat methodemix betekent.
6. De lerarenopleider kan een blauwdruk van een toetsprogramma opstellen.
7. De lerarenopleider kan informatie (resultaten) van verschillende toetsinstrumenten aan elkaar verbinden.
8. De lerarenopleider kan studenten het toetsprogramma uitleggen.
9. De lerarenopleider weet wat 'constructive alignment' is (experts).

Leerdoelen op het niveau van toetsbeleid in context van leven lang leren

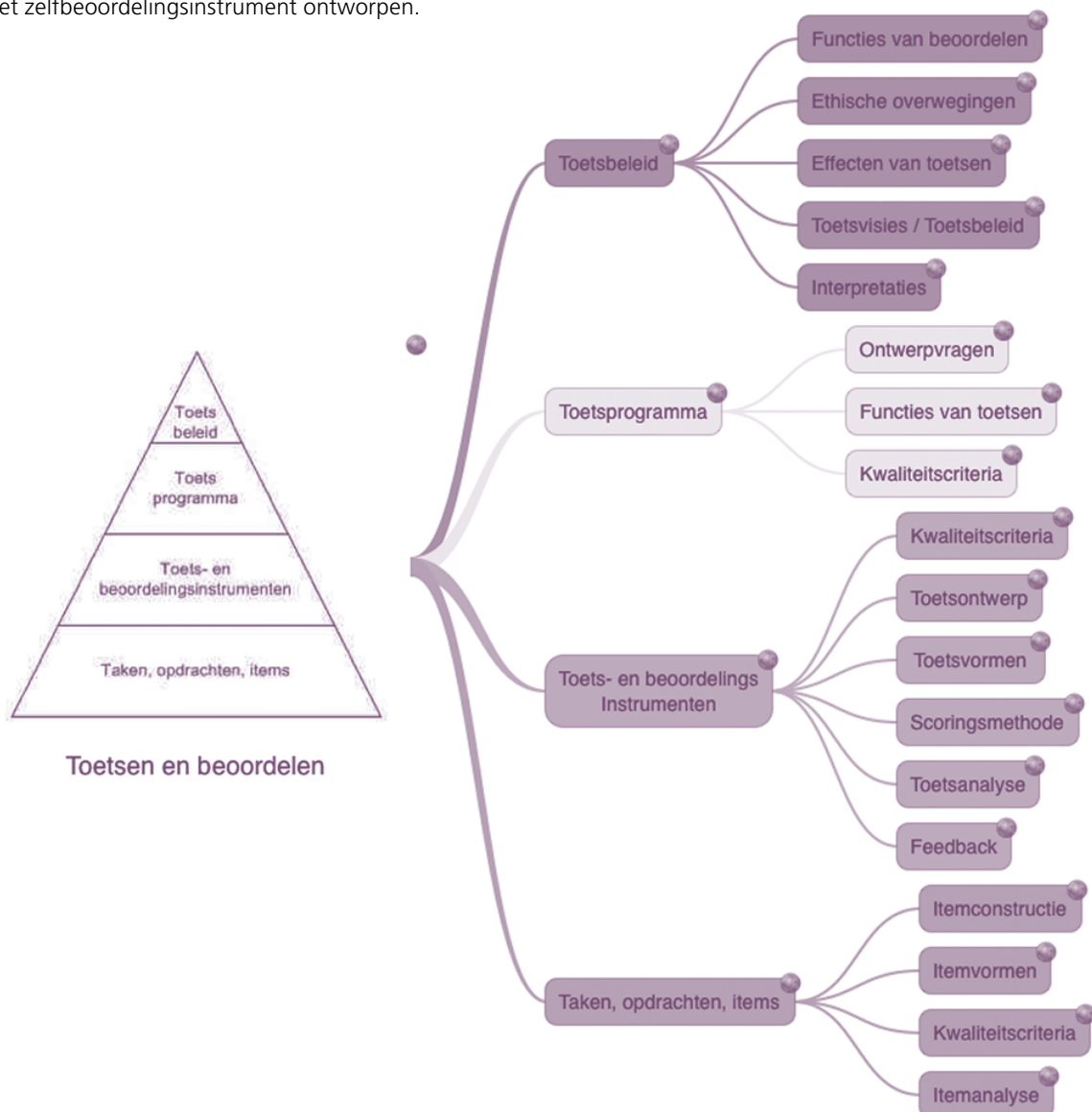
1. De lerarenopleider weet wat externe validering inhoudt.
2. De lerarenopleider weet de visie op toetsing te vertalen in beleid en zijn ontwerp van toetsing (experts).
3. De lerarenopleider weet hoe hij studenten kan betrekken bij het nadenken over toetsvisie.
4. De lerarenopleider kent het onderscheid tussen formeel, non-formeel en informeel beoordelen.
5. De lerarenopleider kent het onderscheid tussen normgericht, criteriumgericht en ipsatief beoordelen.
6. De lerarenopleider weet wat het verschil is tussen toetsen en beoordelen.
7. De lerarenopleider weet hoe hij zelfsturing kan bevorderen in toetsing.
8. De lerarenopleider weet wat duurzaam toetsen betekent.
9. De lerarenopleider kan bij toetsing een coachende rol aannemen om de betrokkenheid, zelfsturing en zelfverantwoordelijkheid van de student te bevorderen (experts).
10. De lerarenopleider weet hoe onethische, illegale en oneerlijke toetspraktijken voorkomen moeten worden.
11. De lerarenopleider weet hoe de kwaliteitsborging vanuit de examencommissies geregeld is.
12. De lerarenopleider kent de achtergronden van toetsen en beoordelen.

Tevens kwamen uit dit literatuuronderzoek verschillende concepten naar voren die gekoppeld zijn aan de leerdoelen. Aan het leerdoel 'De lerarenopleider kan toetsvragen na afname analyseren op psychometrische kwaliteit' is het concept 'itemanalyse' gekoppeld. Deze concepten zijn samengebracht in een conceptmap (Zie

voor een deel van de conceptmap Figuur 9.1; de volledige conceptmap staat op www.bewustentekwaamtoetsen.nl). De kwaliteitspiramide voor eigentijds toetsen en beoordelen, het overzicht van de leerdoelen en de conceptmap hebben samen de basis gevormd voor de ontwikkeling van het zelfbeoordelingsinstrument.

Op basis van de leerdoelen en de conceptmap hebben lerarenopleiders van de vier participerende instellingen de items voor het zelfbeoordelingsinstrument ontworpen.

Figuur 9.1. Een deel van de conceptmap voor het ontwerp van het zelfbeoordelingsinstrument



9.3 Resultaat en conclusie van het project

Het volledige onderzoeksrapport en het zelfbeoordelingsinstrument is terug te vinden op

www.bewustentekwaamtoetsen.nl.

Het onderzoek heeft aangetoond dat het zelfbeoordelingsinstrument een bijdrage kan leveren aan het verhogen van de toetskennis van de lerarenopleiders. Bij dit onderzoek is echter nog wel een aantal kanttekeningen te plaatsen. Het onderzoek geeft geen uitkomsten over de kunde van de lerarenopleiders. In de casustoets die gebruikt werd bij dit onderzoek werd gevraagd wat lerarenopleiders zouden adviseren in een bepaalde situatie. Dit is een kennistoets en de antwoorden willen niet zeggen dat de lerarenopleiders ook daadwerkelijk zo zouden handelen als dat ze hebben aangegeven. Ook is het niet transparant of de docenten bij de evaluatie van het instrument ook de casustoets geëvalueerd hebben. Dit zou de resultaten kunnen hebben beïnvloed.

Het onderzoek is uitgevoerd bij lerarenopleiders. Een vertaling van de resultaten naar alle docenten in het hoger onderwijs lijkt voor de hand te liggen, maar is niet onderzocht. Dit is wel besproken tijdens een van de conferenties die tijdens de projectperiode bezocht zijn. Daarbij werd geconcludeerd dat zowel lerarenopleiders als docenten afhankelijk van hun rol in de opleiding hetzelfde moeten kennen en kunnen, maar dat lerarenopleiders nog meer nadrukkelijk moeten kunnen onderbouwen waarom ze bepaalde keuzes maken in hun toetsontwerp. (zie ook Swennen et al., 2008).

Het project heeft een zelfbeoordelingsinstrument opgeleverd waarmee docenten hun eigen kennis ten aanzien van een breed domein van toetsing kunnen meten. Het instrument is echter ontwikkeld als onderzoeksinstrument, om na te gaan of zo'n instrument zinvol zou kunnen zijn. Daarmee was het doel van dit project niet om een volledig dekkend instrument op te leveren, wat dan ook niet gebeurd is. Het instrument bevat toetsen gespreid over het hele domein, maar is niet volledig dekkend. Dit betekent dat een uitspraak over de kennis van docenten een indicatie is van deze kennis en dan ook uitsluitend formatief, als zelfbeoordeling gebruikt mag worden. De uitslag biedt de gebruiker de mogelijkheid om te reflecteren op zijn kennis..

EAPRIL conferentie in Jyväskylä, Finland.



Hoofdstuk 10

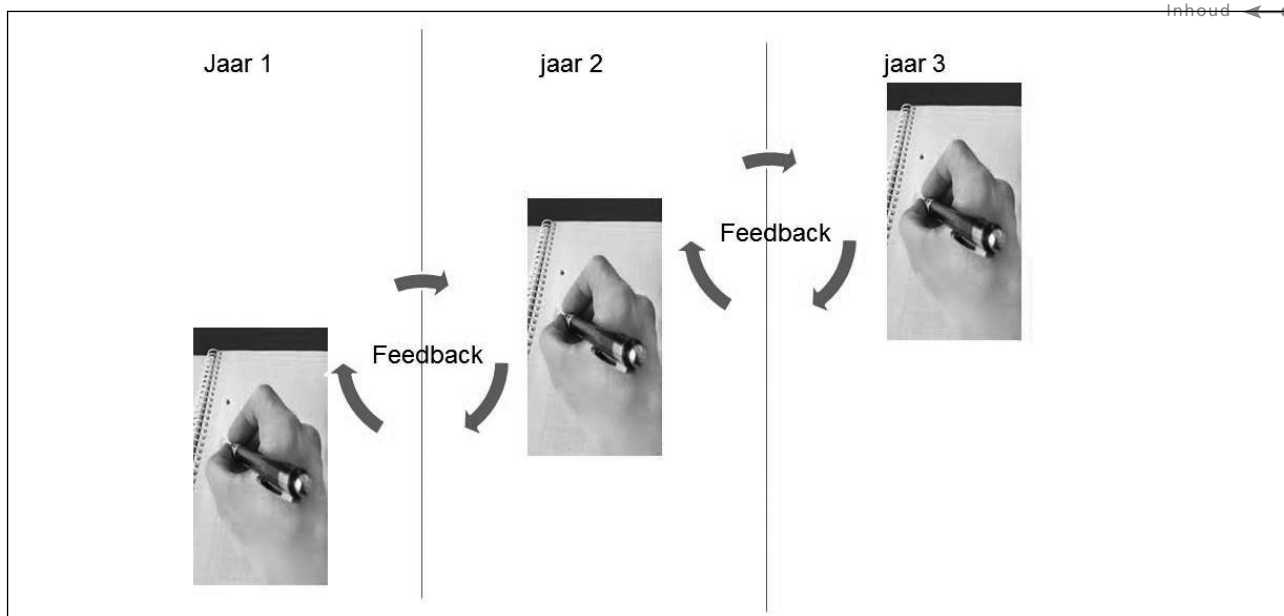
Tot slot

Het doel van lectoraten is professionalisering van docenten, innovatie van het curriculum en het doen van onderzoek. Elk kenniskringlid heeft op basis van eigen onderzoeksvragen praktijkonderzoek uitgevoerd. Daarnaast namen alle leden deel aan het gezamenlijke SURF-onderzoek 'Bewust en bekaam toetsen en beoordelen'.

De verschillende onderzoeken hebben geresulteerd in workshops en trainingen voor collega's en de betrokken kenniskringleden hebben hun onderzoekservaringen direct in de praktijk toegepast. Daarnaast heeft het onderzoek publicaties en presentaties op congressen en conferenties opgeleverd. Het project 'Bewust en bekwaam toetsen' heeft een instrument opgeleverd waarmee docenten hun eigen kennis in kaart kunnen brengen (zie hoofdstuk 9).

Op basis van de resultaten van hun praktijkonderzoek hebben de kenniskringleden hun eigen praktijk aangepast en worden de bevindingen uit hun eigen praktijk vertaald naar curriculumniveau. Het onderzoek van Carel Burghout naar het gebruik van rubrics heeft er in eerste instantie toe geleid dat de rubric door de studenten gezien werd als een zinvol instrument om efficiënt en effectief kwalitatief

betere producten op te leveren. Daarna heeft hij het gebruik van dit instrument verbreed naar een ander vak waarvoor hij zelf verantwoordelijk is. Dit vak had andere doelstellingen en betrof een andere groep studenten. De conclusie was dat het hier ook goed ingezet kon worden, maar dat er wel enkele aanpassingen nodig bleken. Het onderzoek van Monique Maas sluit hier op aan voor weer een vakdidactisch vak. Tegelijkertijd heeft Carel geïnventariseerd bij welke andere vakken in het Talen-domein gebruik gemaakt werd van rubrics. Op dit moment blijkt dat rubrics al wel op meerdere plekken in het curriculum gebruikt worden, maar dat de manier waarop verschilt. Een tweede voorbeeld betreft het onderzoek van Jorik Arts en Mieke Jaspers naar schriftelijke feedback. Zij hebben zich in eerste instantie gericht op het vak waar Jorik docent bij is (actieonderzoek). De resultaten geven een indicatie dat het zinvol is om al eerder in het curriculum bewust om te gaan met het geven van schriftelijke feedback. In de onderzoeksleerlijn, die start in jaar 1 en doorloopt tot het actieonderzoek in jaar 4, wordt nu feedback zodanig ingezet dat de feedback die in het ene jaar gegeven wordt gebruikt moet worden in het daaropvolgende jaar (zie Figuur 10.1).



Figuur 10.1 Het summatief inzetten van formatief toetsen.

Op deze manier ontstaat een betere balans tussen formatief en summatief toetsen. Het afsluitende werkstuk in jaar 1, dat in eerste instantie ontwikkeld is om summatief te toetsen, kan op deze manier effectief gebruikt worden om het leren te stimuleren (Bennett, 2011; Black & Wiliam, 1998; Stiggins, 2005). Om toetsen op deze manier in te zetten is het belangrijk dat de student de juiste informatie krijgt over zijn prestaties door de feedback te koppelen aan de beoordelingscriteria (Sluijsmans, Joosten-ten Brinke, & Van der Vleuten, 2013). Een derde voorbeeld is het onderzoek van Jos Verhoosel naar peer assessment bij groepswerk.

Hierbij was de eerste fase van zijn onderzoek gericht op het vak loopbaanoriëntatie. De inrichting van het groepswerk is op basis van zijn onderzoek aangepast.

De resultaten van alle onderzoeken bleken naar verloop van tijd steeds meer raakvlakken met elkaar te hebben en de betrokken opleiders leren van elkaar hoe ze gebruik kunnen maken van elkaars bevindingen. Voor peer assessment is feedback geven enorm belangrijk en bij het geven van feedback is het werken met rubrics zinvol. Op basis van deze onderzoeken is een vervolgonderzoek geformuleerd op curriculumniveau met als onderzoeksvraag: 'In hoeverre komt formatief beoordelen, waarin aspecten als het geven van feedback

(schriftelijk en mondeling), self en peer assessment, rubrics en het opstellen van criteria belangrijk zijn in het werk van de toekomstige student, bij FLOT in de curricula terug en waar zitten mogelijke verbeterpunten?'. Voordat dit onderzoek verder uitgewerkt kan worden moet vastgesteld worden hoe toetsing nu in het curriculum is verwerkt en moet duidelijk zijn over welke kennis en kunde een docent zou moeten beschikken. In het project Bewust en bekwaam toetsen is voor het beantwoorden van deze laatste vraag een kennisbasis beschreven (zie Tabel 9.1). Hoe kwaliteit van toetsen en beoordelen in het curriculum van de verschillende opleidingen optimaal ingebed moet worden is een onderzoeksopdracht voor de tweede periode van het lectoraat.

10.1 Bijdrage aan de professionalisering van leraren en lerarenopleiders

De professionalisering van leraren op het gebied van toetsen en beoordelen heeft op verschillende manieren plaatsgevonden:

- Professionalisering van opleiders via kenniskring. In totaal hebben 8 lerarenopleiders, een medewerker van O&O en een docent van een vo-school deelgenomen aan de kenniskring van het lectoraat. Zij hebben zich geprofessionaliseerd op twee gebieden: a) inhoudelijk op een specifiek onderwerp van toetsing (te weten formatief toetsen, feedback, rubrics, scripts en checklists, zelfbeoordelingen en peer assessment) en b) kennis en vaardigheden met betrekking tot het doen van praktijkonderzoek.
- Professionalisering van opleiders via pilots. In de eerste paar jaar van het lectoraat zijn scholingsbijeenkomsten verzorgd. Daarna was er begeleiding van pilots op ad hoc basis.
- Professionalisering van opleiders via kennisdeling. Het lectoraat heeft workshops en trainingen georganiseerd rondom toetsing en beoordeling (zie bijlage 1.). Daarnaast hebben alle leden van de kenniskring bijgedragen aan het begeleiden van collega's bij individuele dan wel opleidingsspecifieke vragen op het gebied van toetsing. Deze begeleiding gebeurde 'on the job'. Voorbeelden hiervan zijn gesprekken met alle docenten omgangskunde, instructie over het gebruik van Excel voor analyses bij het talentcluster, instructie over item- en toetsconstructie bij docenten FO en het evalueren van de kwaliteit van rubrics van individuele collega's.

Ook het management is betrokken bij deze professionalisering. Dit is tot uiting gekomen in het samen opstellen van een toetsbeleid en een protocol voor het werken in toetscommissies (onder verantwoordelijkheid van de examencommissie).

Studenten voeren actieonderzoek uit in de bachelor en praktijkonderzoek in de master. Indien toetsing onderwerp is van dit onderzoek kunnen studenten het lectoraat vragen om hen daarbij te begeleiden. In de eerste periode hebben vier studenten van de Masteropleiding Engels daar gebruik van gemaakt: Hans van Bree (afgestudeerd), Martin van der Plas (afgestudeerd), Robert Versteegen en Soraya Khoebal.

10.2 De professionele praktijk

De professionele praktijk van de lerarenopleiding bestaat uit de scholen voor voortgezet onderwijs en middelbaar beroepsonderwijs. Daarnaast behoren tot de relevante beroepspraktijk de collega's van de andere Fontys-instituten en collega's bij andere hogescholen. In de eerste lectoraatsperiode bleek een enorme behoefte aan docentprofessionalisering op het thema toetsing. Tweemaal heeft het lectoraat samen met Onderwijs en Onderzoek een toetsconferentie georganiseerd met Cees van der Vleuten, Dominique Sluijsmans en Janke Cohen-Schotanus als inspirerende sprekers. Deze lezingen hebben bij veel opleidingen gezorgd voor vervolgacties om de kwaliteit van toetsen en beoordelen verder te verhogen.

Fontys heeft zich tot doel gesteld (zie kwaliteitsagenda Fontys) dat elke docent toetsbekwaam is. Dit resultaat moet in 2015 gerealiseerd zijn. Samen met docenten van FHKE PABO Tilburg heeft het lectoraat een bijdrage geleverd aan de BKO-, MKO-, en SKO-handleiding voor Didactiek en toetsing. Deze handleidingen zijn afgestemd op het landelijk programma van eisen (Expertgroep BKE/SKE, 2013).

Het lectoraat heeft verschillende workshops en trainingen (zie bijlage 1) verzorgd om de kennisontwikkeling, kenniscirculatie en kennisimplementatie in de professionele praktijk te stimuleren. Tevens zijn er publicaties verschenen die bij workshops en trainingen gebruikt kunnen worden. Een aantal lezingen en

masterclasses is digitaal te bekijken, waardoor het bereik van de lezingen vergroot wordt.

10.3 Het vervolg

Terugkijkend op deze eerste periode kunnen we vaststellen dat er mooie ontwikkelingen in gang gezet zijn. Het toetsbewust en toetsbekwaam zijn van alle lerarenopleiders, leraren en studenten op alle lagen van de piramide is echter nog niet bereikt. Daar zijn vervolgstappen voor nodig.

In de tweede periode van het lectoraat, dat verder gaat onder de naam 'Kwaliteit van toetsen en beoordelen', zal professionalisering door middel van praktijkonderzoek worden vervolgd. Tevens zal nadrukkelijker gekeken worden in hoeverre onze studenten, de leraren van de toekomst, opgeleid worden tot toetsbekwame docenten. Dit vraagt veel van de opleidingen (zie Tabel 9.1), maar gezien het cruciale belang van het nemen van de juiste beslissingen over leerlingen, is het een uitdaging om hier samen wat moois van te maken.

Literatuur

- Adcroft, A. (2011). The mythology of feedback. *Higher Education Research & Development*, 30, 405-419.
- Allal, L., & Lopez, L. (2005). Formative Assessment of Learning: A Review of Publications in French. *French Literature Review*, 241-264.
- Andrade, H. & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment Research and Evaluation*, 10(3), 1-11.
- Andrade, H. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 13-18.
- Arter, J. & Chappius, J. (2006). *Creating & recognizing quality rubrics*. Portland, OR: Educational Testing Service (ETS).
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18, 5-25.
- Berg, I. van den, Pilot, A., & Admiraal, W. (2005). *Peerassessment als leermiddel. Voorbeelden uit het hoger onderwijs*. Utrecht: IVLOS.
- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing*, 17, 102-118.
- Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5, 7-74.
- Boud, D., & Falchikov, N. (2006). Aligning Assessment with Long-Term Learning. Special Issue: Learning-Oriented Assessment: Principles and Practice. *Assessment & Evaluation In Higher Education*, 31(4), 399-413.
- Boud, D., & Malloy, E. (2013). *Feedback in higher education*. London: Routledge.
- Brennan, R. L. (Eds.) (2006). *Educational Measurement*. 4th edition, American Council on Education.
- Bruno, I., & Santos, L. (2010). Written comments as a form of feedback. *Studies in Educational Evaluation*, 36, 111-120.
- Buhagiar, M. A. (2012). Mathematics student teachers' views on tutor feedback during teaching practice. *European Journal of Teacher Education*, 36, 55-67.
- Burke, K. (2008) *How to Assess Authentic Learning*, 5th edition, Corwin
- Burke, K. (2011). *From Standards to Rubrics in Six Steps: Tools for Assessing Student Learning*. Thousand Oaks, California: Corwin.
- Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research*, 65, 245-281.
- Butler, R. (1988). Enhancing and undermining intrinsic motivation: the effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, 58, 1-14.
- Castelijns, J., Segers, M., & Struyven, K. (2011). *Evalueren om te leren. Toetsen en beoordelen op school*. Bussum: Coutinho.
- Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*, 31, 219-233.
- Cartney, P. (2010). Exploring the use of peerassessment as a vehicle for closing the gap between given feedback and feedback used. *Assessment & Evaluation in Higher Education*, 5 (35), 551-564.
- Cohen, A. (1987). *Student processing of feedback in their composition*. New Jersey: Prentice Hall: Englewood Cliffs.

- Cohen-Schotanus, J. (2010). *Tegenintuïtief*. Oratie, Rijksuniversiteit Groningen, Groningen.
- Commissie Bruijn (2012). *Vreemde ogen dwingen. Eindrapport Commissie externe validering examenkwaliteit hoger beroepsonderwijs*. Den Haag: HBO-raad.
- Davies, W. (2009). Groupwork as a form of assessment: common problems and recommended solutions. *Higher Education*, 58(4), 563-584.
- Dochy, F., Admiraal, W. & Pilot, A. (2003). Peer- en co-assessment als instrument voor diepgaand leren: bevindingen en richtlijnen. *Tijdschrift voor Hoger Onderwijs*, 21(4) 220-229.
- Expertgroep BKE/SKE (2013). Verantwoord toetsen en beslissen in het hoger beroepsonderwijs. Een voorstel voor een programma van eisen voor een basis- en seniorkwalificatie examinering (BKE/SKE). Den Haag: Vereniging Hogescholen.
- Falchikov, N. (2001). *Learning together: Peer tutoring in higher education*. London: RoutledgeFalmer.
- Ferguson, P. (2009). Student perceptions of quality feedback in teacher education. *Assessment & Evaluation in Higher Education*, 36, 51-62.
- Gibbs, G. (2010). Does assessment in open learning support students? *Open Learning: The Journal of Open and Distance Learning*, 25, 163-166.
- Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and teaching in higher education*, 1, 3-31.
- Glover, C., & Brown, E. (2006). Written Feedback for Students: too much, too detailed or too incomprehensible to be effective? *BEE-j*, 7.
- Goodrich, H. (2005). Teaching with rubrics: the Good, the Bad and the Ugly. *College teaching*, 1 (53), 27-31.
- Gratama van Andel, S. (2014). SCALA-Scaffolding Assessment for Learning, Eindrapportage maart 2014. In Projecten programma Toetsing en Toetsgestuurd Leren: SURF foundation.
- Hattie, J. C. (2009). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. London & New York: Routledge, Taylor & Francis Group.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81-112.
- Hernández, R. (2012). Does continuous assessment in higher education support student learning? *Higher Education*, 64, 489-502.
- Hyatt, D. F. (2005). 'Yes, a very good point!': a critical genre analysis of a corpus of feedback commentaries on Master of Education assignments. *Teaching in Higher Education*, 10, 339-353.
- Inspectie van het Onderwijs (2004). *Onderwijsverslag 2002/2003*. Utrecht: Inspectie van het Onderwijs.
- Jaspers, M., & Heijmen-Verstegen, I. (2005). *Toetswijzer competentiegericht begeleiden en beoordelen in het Hoger Onderwijs*. Eindhoven: Fontys Hogescholen, afdeling onderwijs.
- Johnson D. W., & Johnson F. P. (2008). *Groepsdynamica. Theorie en vaardigheden*. (9e editie). Amsterdam: Pearson Education Benelux. ISBN 978-90-430-1490-8.
- Johnson, D. W., & Johnson, R. T. (2005). New Developments in Social Interdependence Theory. *Genetic, Social & General Psychology Monographs*, 131(4), 285-358.
- Johnson, D., Johnson, R., & Smith, K. (2007). The State of Cooperative Learning in Postsecondary and Professional Settings. *Educational Psychology Review*, 19(1), 15-29.
- Jonsson, A. & Svingby, G. (2007). The use of scoring rubrics:

- Reliability, validity and educational consequences. *Educational Research Review*, 2, 130-144.
- Joosten-ten Brinke, D. (2010). *Eigentijds toetsen en beoordelen*. Lectorale rede. Tilburg: Fontys Lerarenopleiding Tilburg. www.fontys.nl/generiek/bronnenbank/sendfile.aspx?id=251356;
- Karau, S. J., & Williams, K. D. (1993). Social Loafing: A Meta-Analytic Review and Theoretical Integration. *Journal Of Personality & Social Psychology*, 65(4), 681-706.
- Karau, S. J., & Williams, K. D. (1995). Social Loafing: Research Findings, Implications, and Future Directions. *Current Directions In Psychological Science (Wiley-Blackwell)*, 4(5), 134-140.
- Lee, I. (2008). Understanding teachers' written feedback practices in Hong Kong secondary classrooms. *Journal of Second Language Writing*, 17, 69-85.
- Lee, I. (2009). Ten mismatches between teachers' beliefs and written feedback practice. *ELT Journal*, 63, 13-22.
- Mabry, L. (1999). Writing to the Rubric: Lingering Effects of Traditional Standardized Testing on Direct Writing Assessment. *Phi Delta Kappan*, 80, 673-679.
- Maiden, B., & Perry, B. (2011). Dealing with Free-Riders in Assessed Group Work: Results from a Study at a UK University. *Assessment & Evaluation In Higher Education*, 36(4), 451-464.
- Mayer, R. E., & Alexander, P. A. (2011). *Handbook of research on learning and instruction* New York and London: Routledge.
- Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation*, 7(25). Retrieved February 8, 2014 from <http://PAREonline.net/getvn.asp?v=7&n=25>.
- Ministerie van OC&W (2008). *Krachtig meesterschap, Kwaliteitsagenda voor het opleiden van leraren 2008-2011*. Den Haag: Ministerie van OC&W.
- Mutch, A. (2003). Exploring the Practice of Feedback to Students. *Active Learning in Higher Education*, 4, 24-38.
- Netherlands Association of Universities of Applied Sciences (2011). Knowledge base English. Retrieved at http://10voordeleraar.nl/documents/kennisbases_master/kb_engels_master.pdf
- Newton, P. M., Wallace, M. J., & McKimm, J. (2012). Improved quality and quantity of written feedback is associated with a structured feedback proforma. *J Educ Eval Health Prof*, 9.
- Nicol, D., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 2 (31), 199-218.
- Nordberg, D. (2008). Group projects: more learning? Less fair? A conundrum in assessing postgraduate business education. *Assessment & Evaluation In Higher Education*, 33(5), 481-492
- NVAO (2004). *Notitie: meta-evaluatie pabo's*. Den Haag: NVAO.
- NVAO (2011). *Beoordelingskaders accreditatiestelsel hoger onderwijs. Opleidingsbeoordeling*. Den Haag: NVAO.
- Orrell, J. (2006). Feedback on learning achievement: rhetoric and reality. *Teaching in Higher Education*, 11, 441-456.
- Orsmond, P., & Merry, S. (2011). Feedback alignment: effective and ineffective links between tutors' and students' understanding of coursework feedback. *Assessment & Evaluation in Higher Education*, 36, 125-136.
- Panadero, E., Alonso Tapia, J. & Huertas, J.A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences*, 22, 806-813. Doi: 10.1016/j.lindif.2012.04.007.

- Perera, J., Lee, N., Win, K., Perera, J., & Wijesuriya, L. (2008). Formative feedback to students: the mismatch between faculty perceptions and student expectations. *Medical Teacher*, 30, 395-399.
- Plake, B. S. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. Mid-Price, M., Handley, K., Millar, J., & O'Donovan, B. (2010). Feedback : all that effort, but what is the effect? *Assessment & Evaluation in Higher Education*, 35, 277-289.
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28, 4-13.
- Reynolds-Keefer, L. (2010). Rubric-referenced assessment in teacher preparation: an opportunity to learn by using. *Practical Assessment, Research and Evaluation*, 15, 1-9.
- Rezai, A.R. & Lovorn, M. (2010) Reliability and validity of rubrics for assessment through writing. *Assessing Writing* 15, 18-39
- Ryan, G. J., Marshall, L. L., Porter, K., & Jia, H. (2007). Peer, Professor and Self-Evaluation of Class Participation. *Active Learning In Higher Education: The Journal Of The Institute For Learning And Teaching*, 8(1), 49-61.
- Sadler, R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education*, 30(2), 175-194.
- Schmidt, H. G. (2012). Hoe actief leren studiesucces beïnvloedt. In H. van Berkel, E. Jansen & A. Bax (Eds.), *Studiesucces bevorderen: het kan en is niet moeilijk*. Den Haag, the Netherlands: Boom Lemma Uitgevers.
- Sluijsmans, D. M. A., van Eldik, S., Joosten-ten Brinke, D., & Jakobs, L. (2013). Bewust en bekwaam toetsen: Wat zouden lerarenopleiders moeten weten over toetsing? *Tijdschrift voor Lerarenopleiders*, 34(3), 27-42.
- Sluijsmans, D.M.A. (2002). *Student involvement in assessment. The training of peer assessmentskills*. Heerlen: Open Universiteit.
- Sluijsmans, D.M.A., Prins, F. (2006). A conceptual framework for integrating peerassessment in teacher education. *Studies in Educational Evaluation*. 32, 6-22.
- Stiggins, R. (2005). From formative assessment to assessment FOR learning: A path to success in standards-based schools. *Phi Delta Kappa*, 87, 324-328.
- Straetmans, G. J. J. M. (2006). *Bekwaam beoordelen en beslissen* (lectorale rede). Deventer: Saxion Hogescholen.
- Tomlinson, C.A. (2003). Fulfilling the promise of the differentiated classroom: Strategies and tools for responsive teaching. Alexandria, VA: Association for Supervision and Curriculum Development.
- Topping, K. J. (2009). *Peer Assessment. Theory Into Practice*, 48(1), 20-27.
- Tractenberg, R. E., Umans, J. G., & McCarter, R. J. (2009). A Mastery Rubric: guiding curriculum design, admissions and development of course objectives. *Assessment & Evaluation in Higher Education*, 35, 15-32.
- Van Berkel, A., Sluijsmans, D., & Joosten-ten Brinke, D. (in voorbereiding). Toetsdeskundigheid.
- Van Berkel, H., Bax, A., & Joosten-ten Brinke, D. (2014). Het toetsproces ontleed. In H. Van Berkel, A. Bax, & D. Joosten-ten Brinke, *Toetsen in het hoger onderwijs* (pp. 1-11). Houten: Bohn, Stafleu Van Lochem.
- Van Meeuwen, L. W., Brand-Gruwel, S., Kirschner, P. A., De Bock, J. J. P. R., & Van Merriënboer, J. J. G. (2012). Vertaling en validatie van twee vragenlijsten: Self-efficacy en self-directed learning. Paper presented at the Onderwijs Research Dagen, Wageningen.

- Walker, M. (2007). Improved learning through improved feedback on assignments. In: The open university.
- Walker, M. (2009). An investigation into written comments on assignments: do students find them usable? *Assessment & Evaluation in Higher Education*, 34, 67-78.
- Weaver, M. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment & Evaluation in Higher Education*, 31, 379-394.
- William, D. (2011) *What is Assessment for Learning?* Studies in Educational Evaluation 37, 3-14
- Wilson, N. S. (2008). Teachers expanding pedagogical content knowledge: Learning about formative assessment together. *Journal of In-Service Education*, 34, 283-298.
- Wood, D. (2009). A scaffolded approach to developing students' skills and confidence to participate in self and peer assessment. *ATN Assessment Conference* (pp. 374-384). RMIT University.
- Wulf, G., Raupach, M., & Pfeiffer, F. (2005). Self-controlled observational practice enhances learning. *Research Quarterly for Exercise and Sport*, 76, 107-111.
- Xiao, Y. (2010, January 1). The Effects of Training in Peer Assessment on University Students' Writing Performance and Peer Assessment Quality in an Online Environment. ProQuest LLC.
- Zundert, M. van. (2012). *Conditions of peer assessment for complex learning*. Maastricht: Universitaire Pers Maastricht.
- Zundert, M. van., Sluijsmans, D., Merrienboer, J.J.G. van. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and instruction*, 20, 270-279.

CV van de lectoraatsleden



Desirée Joosten-ten Brinke

Dr. Desirée Joosten-ten Brinke studeerde Toegepaste Onderwijskunde aan de Universiteit Twente en is afgestudeerd bij de vakgroep Onderwijskundige Meetmethoden en Data-analyse. Vanaf 1990 is zij werkzaam als toetsdeskundige/onderzoeker bij de Open universiteit. Zij promoveerde daar in 2008 op het proefschrift getiteld 'Assessment of prior learning'. Vanaf 2010 is zij tevens lector Eigentijds toetsen en beoordelen bij Fontys lerarenopleiding Tilburg. Haar onderzoeks- en onderwijsfocus is gericht op kwaliteit van toetsen en beoordelen, digitaal toetsen, EVC en formatief beoordelen. Zij begeleidt docenten en promovendi bij hun onderzoek naar deze thema's en verzorgt trainingen voor docenten en leden van examencommissies.

Desirée was projectleider van het SURF-project 'Bewust en bekwaam toetsen en beoordelen', en voorzitter van de kwaliteitscommissie voor de invoering van de landelijke kennistoetsen voor de lerarenopleidingen (10vdl). Zij is lid van visitatiecommissies in opdracht van de nqa en hoofdredacteur van Examens, tijdschrift voor de toetspraktijk.



Jorik Arts

Jorik Arts studeerde Medische Biologie aan de Vrije Universiteit in Amsterdam. Na voltooiing van deze opleiding volgde een promotieonderzoek aan de vakgroep moleculaire microbiologie van de Universiteit Utrecht. Daar promoveerde hij in 2007. Sinds 2007 is hij werkzaam bij de Fontys lerarenopleiding in Tilburg. Hij is daar vanaf de start van de aanstelling betrokken geweest bij het thema toetsen. Hij schreef mee aan de toets nieuwe opleiding voor de aanvraag van een masteropleiding van de lerarenopleiding Biologie en aan de accreditierapporten voor zowel de Bachelor- als Masteropleiding. Hij is sinds 2008 voorzitter van de toetscommissie Biologie en sinds 2013 voorzitter van de toetscommissie Beta-plus (lerarenopleidingen natuurkunde, scheikunde, techniek, biologie, GZW en omgangskunde). Daarnaast is hij sinds 2010 lid van de kenniskring Eigentijds toetsen en beoordelen bij Fontys lerarenopleiding Tilburg.



Carel Burghout

Drs Carel Burghout studeerde in 1976 af aan de Stichting Opleiding Leraren, Utrecht (2e graad Engels en Geschiedenis) en in 1978 aan de Universiteit van Utrecht (Engels en 1e-graads leraar). Hij heeft vanaf 1976 in het onderwijs gewerkt, sinds 1990 is hij docent en lerarenopleider bij Fontys Lerarenopleidingen Tilburg. In die hoedanigheid was hij betrokken bij diverse landelijke projecten, zoals het ontwikkelen van een geschikt portfolio voor het beoordelen van competenties en het kijkkader ervoor voor de lerarenopleidingen, de landelijke kennisbasis Bachelor en de landelijke kennisbasis Master en recent bij LOVM, de ijking van toetsing bij en door de landelijke vak Master opleidingen.

Sinds 1998 is hij een actief lid van het uitvoerende orgaan voor toetsing (Testing Evaluation and Assessment) binnen IATEFL (the International Association of Teachers of English as a Foreign Language) onder andere drie jaar als coördinator. Voor IATEFL spreekt Carel regelmatig over toetsing, met name over Assessment for Learning, en over literatuuronderwijs.

In de Kenniskring Eigentijds toetsen en beoordelen heeft hij meegewerkt aan het SURF project 'Bewust en Bekwaam Beoordelen'. De focus van zijn onderzoek binnen de kenniskring ligt op de invloed en effecten van het inzetten van rubrics als instrument voor zowel formatief als summatief assessment. Een neveneffect van zijn onderzoek en betrokkenheid bij de Kenniskring is dat hij zowel binnen FLOT als daarbuiten regelmatig workshops over rubrics geeft.



Mieke Jaspers

Mieke Jaspers MA studeerde o.a. Onderwijskunde en rondde in 1999 de opleiding Management of learning and development af aan de Universiteit van Tilburg, Business School Tias/Nimbas.

Ruim 35 jaren is zij actief werkzaam op het terrein van toetsing en examinering zowel in het reguliere onderwijs als de marktgerichte sector in het VMBO/ MBO/HBO Bachelor/Master.

Vanaf 2002 is zij werkzaam als senior consultant bij de Dienst Onderwijs en Onderzoek, Fontys Hogescholen.

Vanaf 2010 is zij tevens lid van het lectoraat Eigentijds toetsen en beoordelen bij Fontys leraren opleiding Tilburg waarin zij samen met Jorik Arts onderzoek doet naar het effect van schriftelijk feedback op leren en opleiden.

Regelmatig wordt zij gevraagd om te spreken op landelijke congressen rond de thema's toetsing en examinering en de rol van examencommissies in het Hoger Onderwijs. Daarnaast is zij werkzaam als externe examinerator en auditor bij instellingen in het hoger onderwijs. Ook publiceert zij regelmatig over het thema toetsing en examinering.



Monique Maas

Drs. Monique Maas studeerde Engelse taal en literatuur aan de Universiteit Utrecht.

Via een Harting Scholarship studeerde zij een jaar aan de University of Newcastle upon Tyne. Met een Fulbright beurs verbleef zij 6 weken in de U.S.A. Zij volgde aansluitend de universitaire lerarenopleiding te Utrecht (IVLOS). Sinds haar afstuderen in 1997 heeft zij aan de lerarenopleidingen van Tilburg (FLOT) en de Hogeschool Utrecht gewerkt. Sinds 2011

neemt zij in Tilburg stagebegeleiding voor haar rekening neemt en doceert zij vakdidactiek, literatuur en geschiedenis van de Engelse taal. Tevens is zij lid van de kenniskring van het lectoraat Eigentijds Toetsen en Beoordelen onder leiding van Dr. Desirée Joosten-ten Brinke.

Haar onderzoeksfocus bij het praktijkonderzoek is gericht op het verbeteren van de cursussen vakdidactiek voor de eerstejaars deeltijd studenten aan de lerarenopleiding Engels.



Jos C.M. Verhoosel

Drs. Jos C.M. Verhoosel studeerde Experimentele Natuurkunde aan de Rijksuniversiteit van Utrecht. Na zijn afstuderen in 1978 werd hij werkzaam als docent wis- en natuurkunde op een lyceum. Vanaf 1985 is hij werkzaam bij Fontys, eerst bij de PTH, later bij de Fontys lerarenopleiding in Tilburg als wiskundedocent in het bèta-cluster. Als zodanig begeleidt hij ook masterstudenten bij het uitvoeren van praktijkonderzoek. Sinds 2010 is hij lid van het lectoraat Eigentijds Toetsen en beoordelen. Zijn onderzoekfocus is hierbij vooral gericht op peer-assessment en het beoordelen van groepswerk.



Wim Lansu

Drs. Wim Lansu studeerde in 1989 af aan de lerarenopleiding van de Hogeschool Katholieke Leergangen Tilburg (2e graad Geschiedenis en Nederlands) en in 1992 aan de Rijks Universiteit Utrecht (doctoraal politieke geschiedenis) en aan de HKLT (1e graads geschiedenis). Hij heeft vanaf 1991 in het voortgezet onderwijs gewerkt en sinds 1993 als lerarenopleider bij de Fontys Lerarenopleiding Tilburg. Vanuit deze functie was hij betrokken bij verschillende landelijke projecten, zoals de kennisbases voor de lerarenopleidingen geschiedenis (bachelor en master); de landelijke kennistoets voor bachelor geschiedenis en bij het landelijk vakoverleg voor de masteropleidingen geschiedenis. In de Kenniskring Eigentijds toetsen en beoordelen heeft hij meegewerkt aan het SURF project 'Bewust en Bekwaam Beoordelen'. In het verlengde hiervan geeft hij binnen en buiten FLOT regelmatig workshops over toetsconstructie.

**Suzanne Kratsborn**

Suzanne Kratsborn studeerde Schoolmuziek en Dwarsfluit D.M. aan het Conservatorium en Pabo. Sinds 2000 is zij docent bij FHKE Pabo Tilburg. Zij is lid van de toetscommissie van FHKE. Van 2012-2014 nam zij deel aan een professionaliseringstraject van dienst Onderwijs & Onderzoek. Hierin richtte zij zich voornamelijk op (de kwaliteit van) toetsing. Van 2010-2012 nam zij deel aan het lectoraat Eigentijds toetsen en beoordelen. Voor de Master Leren & Innoveren deed ze onderzoek naar beoordelingsvaardigheden die noodzakelijk zijn voor het uitvoeren van peer assessment: De titel van het onderzoek is Trainen van Beoordelingsvaardigheden: Een aanbeveling voor Fontys PABO Tilburg voor het effectief en efficiënt gebruik van peerassessments door het trainen van beoordelingsvaardigheden.

Bijlage 1

Overzicht van workshops, presentaties en publicaties in 2010-2014

Workshops en trainingen

Oomens, A. & Joosten-ten Brinke, D. (2014, 22 januari). Kwaliteit van toetsprogramma's. Workshop tijdens de Week van de toetsvragen. Tilburg: Fontys lerarenopleiding Tilburg.

Verhoosel, J. & Joosten-ten Brinke, D. (2014, 22 januari). Het beoordelen van groepswerk. Workshop tijdens de Week van de toetsvragen. Tilburg: Fontys lerarenopleiding Tilburg.

Burghout, C. & Maas, M. (2014, 21 januari). Het werken met rubrics, scripts en checklists. Workshop tijdens de Week van de toetsvragen. Tilburg: Fontys lerarenopleiding Tilburg.

Langendonck, K. & Joosten-ten Brinke, D. (2014, 21 januari). Het construeren van een goede toetsvraag. Workshop tijdens de Week van de toetsvragen. Tilburg: Fontys lerarenopleiding Tilburg.

Joosten-ten Brinke, D. (2014, 20 januari). Het analyseren van productbeoordelingen. Workshop tijdens de Week van de toetsvragen. Tilburg: Fontys lerarenopleiding Tilburg.

Arts, J. & Jaspers, M. (2014, 20 januari). Feedback, feed-up en feed-forward. Workshop tijdens de Week van de toetsvragen. Tilburg: Fontys lerarenopleiding Tilburg.

Joosten-ten Brinke, D. (2013). Validiteit van toetsing. Discussie bij gamma. Tilburg: Fontys. Workshops bij de innovatieochtend

Joosten-ten Brinke, D., Burghout, C., Arts, J., Maas, M., & Verhoosel, J. (2013, 31 oktober). Praktijkonderzoek naar je eigen beoordelingspraktijk. Workshop bij de Fontys inspiratiedag. Eindhoven: Fontys.

Lansu, W., Arts, J., Maas, M., & Joosten-ten Brinke, D. (2012). Hoezo, een goed proefwerk? Workshop bij de docentendag. Tilburg: FLOT.

Lansu, W. & Joosten-ten Brinke, D. (2012). Beoordelen van meerkeuzevragen. 11 mei 2012. Presentatie tijdens de Week van de toetsvragen. Tilburg: Fontys lerarenopleiding Tilburg.

Joosten-ten Brinke, D. & Burghout, C. (2012). Assessment for Learning. 10 mei 2012. Presentatie tijdens de Week van de toetsvragen. Tilburg: Fontys lerarenopleiding Tilburg.

Joosten-ten Brinke, D. (2012). Toetsconstructie. 9 mei 2012. Presentatie tijdens de Week van de toetsvragen. Tilburg: Fontys lerarenopleiding Tilburg.

Kratsborn, S. & Joosten-ten Brinke, D. (2012). Peer assessment. 9 mei 2012. Presentatie tijdens de Week van de toetsvragen. Tilburg: Fontys lerarenopleiding Tilburg.

Burghout, C., & Joosten-ten Brinke, D. (2012). Hoe maak je een proefwerk? Workshop bij de meesterschapsdagen voor 1e jaars, 9 februari 2012. Tilburg: FLOT.

- Joosten-ten Brinke, D., & Oomens, A. (2012). Het belang van toetsmatrjjs. Leraar24. <http://www.leraar24.nl/video/3505>
- Joosten-ten Brinke, D. (2011, 1 februari). (on)mogelijkheden van digitale toetsen. Lunchlezing, FLOT, Tilburg. <http://hdl.handle.net/1820/3812> Workshop of presentatie
- Joosten-ten Brinke, D. (2014, 8 januari). Scholingsdag toetsing voor docenten van de opleiding Duits. Sittard: FLOS.
- Sluijsmans, D. M. A. & Joosten-ten Brinke, D. (28/913-1/5/2014). Leergang toetsdeskundigheid docenten SVOPL. Heerlen: SVOPL.
- Joosten-ten Brinke, D. (2013, 17 december). Digitaal toetsen. De rol van ICT bij toetsen. Webinar voor de VOR-ICT. Eindhoven: Fontys Hogescholen. <http://fontys.mediamission.nl/Mediasite/Play/efe26b3ff7b14ba18b124551b711ce8c1d?catalog=28df4668-ce8a-4a4b-b5f9-d6f6fa7b0d66>
- Joosten-ten Brinke, D. (2013, 6 november). Toetsdeskundigheid in examencommissies. Training voor leden examencommissies Hogeschool Leiden. Leiden: Vereniging Hogescholen.
- Joosten-ten Brinke, D., Burghout, C., Arts, J., Maas, M., & Verhoosel, J. (2013, 31 oktober). Praktijkonderzoek naar je eigen beoordelingspraktijk. Workshop bij de Fontys inspiratiedag. Eindhoven: Fontys.
- Joosten-ten Brinke, D. (2013, 2 december). De intuïtie voorbij. De toetsbewuste onderwijspraktijk. Presentatie bij het Fontysberaad. Heeze: Fontys.
- Joosten-ten Brinke, D. & Burghout, C. (2013, 20 november). Judging of complex tasks with and by students. Masterclass Assessment and Judgement. Maastricht: UM.
- Sluijsmans, D. M. A. & Joosten-ten Brinke, D. (2013, 23 oktober). Formative assessment. What is it and how can it be used in higher education to shape students' learning? Masterclass Assessment and Judgement. Maastricht: UM.
- Joosten-ten Brinke, D. (2013, 29 oktober). Werken met toetsmatrjzen. Workshop op studiedag. Sittard: FLOS.
- Joosten-ten Brinke, D. (2013, 16 oktober). Studiesucces door toetsdeskundigheid docenten. Lezing bij HvA Toetsdebat. Amsterdam: Hogeschool van Amsterdam. <http://hva.mediamission.nl/Mediasite/Play/f3d7e59a602d4d21bb247593ff94d5d11d>
- Joosten-ten Brinke, D. (2013, 10 september). Van brugklas naar eindexamen een continue ontwikkeling. Studiedag Odulphus College. Tilburg: Odulphus.
- Joosten-ten Brinke, D. (2013, 3 september). Toets- en itemanalyse. Feedbacksessie Surf-project. Gouda: Driestar Hogeschool.
- Joosten-ten Brinke, D. (2013, 17 juni). Toetsing: waarom en hoe?. Presentatie bij het management van de opleiding Marketing en Management (FHMM). Eindhoven: Fontys.
- Joosten-ten Brinke, D. (2013, 14 juni). Training leden toetscommissie Toegepaste psychologie. Amsterdam: Hogeschool van Amsterdam.
- Joosten-ten Brinke, D. (2013, 7 juni). Toetsing: op weg waar naar toe? Presentatie voor de AOB. Utrecht: AOB.
- Joosten-ten Brinke, D., (2013, 20 en 21 maart). Toetsdeskundigheid in examencommissies. Training voor leden examencommissies. Utrecht: Hbo-raad.
- Joosten-ten Brinke, D. (2013, 7 maart). Kwaliteit van toetsing - borging eindniveau. Scholingsdag bij Stenden, Drachten.
- Joosten-ten Brinke, D. (2013, 28 februari). Assessment literacy. Lecture at the ICO-course 'Multiple perspectives on assessment, Maastricht.
- Joosten-ten Brinke, D., (2013, 1 februari). Balanceren tussen formatief en summatief toetsen. [Balancing formative and summative assessment]. Presentation at the masterclass

- 'Hoe balancer je formatief en summatief toetsen?' [How to balance formative and summative assessment?], Heerlen, The Netherlands.
- Joosten-ten Brinke, D. (2012). Toets- en itemanalyse. Workshop bij een studiedag voor docenten van De Nieuwste Pabo. 9 november 2012. Sittard: Zuyd Hogeschool en Fontys Hogescholen.
- Joosten-ten Brinke, D., & Moerkerke, G. (2012). Toetsdeskundigheid in examencommissies. Training voor leden examencommissies. 30 november en 1 december 2012. Utrecht: Hbo-raad.
- Joosten-ten Brinke, D. (2012, 3 oktober). Kwaliteit van toetsing. Presentatie gegeven tijdens de scholingsdag over kwaliteit van toetsing, Nijmegen, Nederland: Hogeschool van Arnhem en Nijmegen. <http://hdl.handle.net/1820/4761>.
- Joosten-ten Brinke, D. (2012). Toetsconstructie. 29 mei 2012. Presentatie op de scholingsmiddag van Mill Hill college. Goirle: Mill Hill college.
- Joosten-ten Brinke, D. (2012). Toetsdeskundigheid in examencommissies. Training voor leden examencommissies. 25 en 26 april 2012. Utrecht: Hbo-raad.
- Joosten-ten Brinke, D. (2012). Trends en dilemma's bij toetsing in het Hoger onderwijs. 26 maart 2012. Presentatie bij de Onderwijs Inspectie. Utrecht: Onderwijs Inspectie. <http://hdl.handle.net/1820/4755>
- Joosten-ten Brinke, D. (2012, 19 januari). Standaard setting - proeve van bekwaamheid. Presentatie voor docenten SPL, Amersfoort, Nederland: Stichting Praktijk Leren. <http://hdl.handle.net/1820/4760>.
- Joosten-ten Brinke, D. (2012). Scholingsdagen Objectief beoordelen van vaardigheden. 17 januari 2012. Best: Heerbeek College.
- Joosten-ten Brinke, D. & Kratsborn, S. (2012, 13 januari). Self-assessment en peer-assessment. Lezing, Hogeschool de Kempel, Helmond.
- Joosten-ten Brinke, D. (2011, 25 oktober). Kwaliteit van toetsen en beoordelen. Lezing verzorgd tijdens studiedagen toetsing, Rotterdam: Hogeschool Rotterdam. <http://hdl.handle.net/1820/3777>.
- Joosten-ten Brinke, D. (2011, 27 september). Kwaliteit van toetsen en beoordelen. Lezing verzorgd tijdens scholingsdagen Archimedes, Utrecht: Hogeschool Utrecht. <http://hdl.handle.net/1820/3778>.
- Joosten-ten Brinke, D., & Jaspers, M. (2011, 6 april). Toetsdeskundigheid in examencommissies. Werksessie verzorgd tijdens de Dag van de examencommissies. Utrecht: HBO-raad. <http://hdl.handle.net/1820/3779>.
- Joosten-ten Brinke, D. (2011, 18 maart). Digitaal toetsen. Lunchlezing, FLOS, Tilburg. <http://hdl.handle.net/1820/3780>.
- Verhoosel, J. (2011). Random assessed Peer assessment. Presentatie bij de ElWier bijeenkomst. Hogeschool Arnhem en Nijmegen.
- Joosten-ten Brinke, D. (2011). Scholingsdagen Kwaliteit van toetsen en beoordelen. 27 januari en 21 februari 2011. Eindhoven: Stedelijk College Eindhoven.
- Joosten-ten Brinke, D. (2010, 27 mei). Kwaliteit van toetsen. Scholingsdag voor IEMES. Tilburg: FLOT.

Publicaties

- Van Berkel, H., Bax, A., & Joosten-ten Brinke, D. (2014). Toetsen in het hoger onderwijs. (3e geheel herziene druk). Houten: Bohn, Stafleu Van Lochem.
- Van Berkel, H., Bax, A., & Joosten-ten Brinke, D. (2014). Het toetsproces ontleed. In H. Van Berkel, A. Bax, & D. Joosten-ten Brinke, Toetsen in het hoger onderwijs. (pp. 1-11). Houten: Bohn, Stafleu Van Lochem.
- Joosten-ten Brinke, D. & Sluijsmans, D.(2014). Formatief toetsen. In H. Van Berkel, A. Bax, & D. Joosten-ten Brinke, Toetsen in het hoger onderwijs. (pp. 52-58). Bohn, Stafleu Van Lochem.
- Van den Bos, P., Burghout, C., & Joosten-ten Brinke, D.(2014). Toetsen met rubrics. In H. Van Berkel, A. Bax, & D. Joosten-ten Brinke, Toetsen in het hoger onderwijs. (pp. 134-142). Houten: Bohn, Stafleu Van Lochem.
- Cortenbach, T., Janssen-Seezink, A., Teune, P., Joosten-ten Brinke D., & Jaspers, M. (2014). Handleiding Fontys Assessment didactiek/toetsing Medior kwalificatie Onderwijs (MKO). Tilburg/Eindhoven: Fontys.
- Cortenbach, T., Janssen-Seezink, A., Teune, P., Joosten-ten Brinke D., & Jaspers, M. (2014). Handleiding Fontys Assessment didactiek/toetsing Senior kwalificatie Onderwijs (SKO). Tilburg/ Eindhoven: Fontys.
- Klarus, R. & Joosten-ten Brinke, D. (2014). Toetsen met EVC. In H. Van Berkel, A. Bax, & D. Joosten-ten Brinke, Toetsen in het hoger onderwijs. (pp. 142-150). Houten: Bohn, Stafleu Van Lochem.
- Joosten-ten Brinke, D. (2012). Gezien en gelezen. Psychometrics in practice at RCEC. Examens, Tijdschrift voor de toetspraktijk, 9(3), 36. <http://hdl.handle.net/1820/4758>
- Joosten-ten Brinke, D. (2012). Gezien en gelezen. Het platform digitaal toetsen. Examens, Tijdschrift voor de toetspraktijk, 9(2), 36. <http://hdl.handle.net/1820/4757>.
- Joosten-ten Brinke, D. (2012). Toetsing en beoordelen in het onderwijs: een continu onderzoek naar de ontwikkeling van leerlingen of studenten. In S. Bolhuis & Q. Kools (samenstellers), Praktijkonderzoek als professionele leerstrategie in onderwijs en opleiding. (pp. 49-68). Tilburg: FLOT. http://www.hbo-kennisbank.nl/nl/page/hborecord.view/?uploadId=fontys_didlmods%3Aoi%3Arepository.samenmaken.nl%3Ampid%3A29826
- Van Berkel, H., & Joosten-ten Brinke, D.(2012). Pieter Drenth. In H. Van Berkel & A. De Knecht-Van Eekelen (samenstellers), Kopstukken uit de Nederlandse examenwereld. (pp. 57-62). Neer: Kloosterhof.
- Joosten-ten Brinke, D. & De Jongh, A. (2012). Cees van der Vleuten. In H. Van Berkel & A. De Knecht-Van Eekelen (samenstellers), Kopstukken uit de Nederlandse examenwereld. (pp. 52-56). Neer: Kloosterhof.
- Molkenboer, H. & Joosten-ten Brinke, D. (2012). Wim van der Linden. In H. Van Berkel & A. De Knecht-Van Eekelen (samenstellers), Kopstukken uit de Nederlandse examenwereld. (pp. 25-29). Neer: Kloosterhof.
- Van Berkel, H. & Joosten-ten Brinke, D. (2012). Wim Hofstee. In H. Van Berkel & A. De Knecht-Van Eekelen (samenstellers), Kopstukken uit de Nederlandse examenwereld. (pp. 20-24). Neer: Kloosterhof.
- Joosten-ten Brinke, D., Lansu, W., Van Leuven, A., & Sluijsmans, D. (2012). Gewogen en ...? De landelijke kennistoets geschiedenis. KLEIO, 53(1), 32-34.

- Kools, Q. & Joosten-ten Brinke, D. (2011) Lectoraten op dreef. Vector Magazine, 17, 17-18.
- Sluijsmans, D. M. A., & Joosten-ten Brinke, D. (2011). Wanneer heeft iemand iets geleerd? In: M. Lunenberg en J. Dengerink (Eds), Kennisbasis lerarenopleiders. Eindhoven: Velon. <http://dspace.ou.nl/handle/1820/3810>
- Van Berkel, H., & Joosten-ten Brinke, D. (2011). Kopstukken uit de examenwereld. Pieter Drenth, de man van wetenschap en bestuur. Examens, Tijdschrift voor de toetspraktijk, 4, 28-32. <http://hdl.handle.net/1820/3923>.
- Joosten-ten Brinke, D. & De Jongh, A. (2011). Kopstukken uit de examenwereld. Cees van der Vleuten, voorstander van longitudinaal toetsen. Examens, Tijdschrift voor de toetspraktijk, 3, 29-32. <http://hdl.handle.net/1820/3783>.
- Joosten-ten Brinke, D. (2011). Gezien en gelezen. Toetsen op school. Examens, Tijdschrift voor de toetspraktijk, 3, 33. <http://hdl.handle.net/1820/3782>.
- Joosten-ten Brinke, D. (2010). Gezien en gelezen. Competent erkend? Examens, Tijdschrift voor de toetspraktijk, 2, 29-30. <http://hdl.handle.net/1820/3164>
- Molkenboer, H. & Joosten-ten Brinke, D. (2010). Kopstukken. Wim van der Linden, itembanking centraal. Examens, Tijdschrift voor de toetspraktijk, 2, 25-28. <http://hdl.handle.net/1820/3155>
- Van Berkel, H. & Joosten-ten Brinke, D. (2010). Kopstukken. Wim Hofstee. Examens, Tijdschrift voor de toetspraktijk, 1, 21-25. <http://hdl.handle.net/1820/3156>
- Joosten-ten Brinke, D., Arts, J., Jaspers, M., Burghout, C., Maas, M. & Verhoosel, J. (2014). Praktijkonderzoek naar de eigen beoordelingspraktijk. In Q. Kools & E. Klatter, Inspiratiedag. (pp. 103-115). Eindhoven/Tilburg: Fontys. <http://goo.gl/Z21vN7>
- Arts, J., Jaspers, M., & Joosten-ten Brinke, D. (submitted). Written feedback as a form of feedback: So much to gain.
- Sluijsmans, D. M. A., Joosten-ten Brinke, D., & Van der Vleuten, C. P. M. (2013). Toetsen met leerwaarde. Een reviewstudie naar effectieve kenmerken van formatief toetsen. Den Haag: NWO.
- Sluijsmans, D. M. A., van Eldik, S., Joosten-ten Brinke, D., & Jakobs, L. (2013). Bewust en bekwaam toetsen: Wat zouden lerarenopleiders moeten weten over toetsing? Tijdschrift voor Lerarenopleiders, 34(3), 27-42.
- Joosten-ten Brinke, D., & Sluijsmans, D. (2012). Tijd voor toetskwaliteit: het borgen van toetsdeskundigheid van examencommissies. TH&MA, 19(4), 16-21. <http://hdl.handle.net/1820/4759>.
- Joosten-ten Brinke, D., Lansu, W., Van Leuven, A., & Sluijsmans, D. (2012). Gewogen en ...? De landelijke kennistoets geschiedenis. KLEIO, 53(1), 32-34.
- Joosten-ten Brinke, D. (2011). Eigentijds toetsen en beoordelen. Lectorale rede. Tilburg: Fontys Lerarenopleiding Tilburg. www.fontys.nl/generiek/bronnenbank/sendfile.aspx?id=251356
- Joosten-ten Brinke, D., Blom, S., & Meusen-Beekman, K. M. (2011). Hoe goed ken je jezelf? Zelfbeoordeling en zelfreflectie in het onderwijs. In J. Castelijns, M. Segers, & K. Struyven (Eds), Evalueren om te leren. (pp. 121-135). Bussum: Coutinho.

Sluijsmans, D. M. A., & Joosten-ten Brinke, D. (2011). Wanneer heeft iemand iets geleerd? In: M. Lunenberg en J. Dengerink (Eds), Kennisbasis lerarenopleiders. Eindhoven: Velon. <http://dspace.ou.nl/handle/1820/3810>

Presentaties en workshops op congressen en conferenties

Burghout, C. & Maas, M. (2014, Januari). Workshop Rubrics tijdens de innovatieochtend .Fontys Lerarenopleiding Tilburg

Burghout, C (2014, April) . Classroom Testing and Assessment for Learning. IATEFL Pre-Conference Event, Harrogate, UK.

Joosten-ten Brinke, D., Van Eldik, S., & Florack, A. (2013, 12 november). Bewust en bekwaam toetsen: een instrument voor zelfbeoordeling en zelfstudie. Presentatie bij de SURF Onderwijsdagen. Rotterdam: SURF.

Burghout, C. (2013, June). Assessment for Learning for Young Learners. NAFLE Conference, Beijing,

Burghout, C. (2013, November). Workshop rubrics voor het Talenteam Fontys Lerarenopleiding Tilburg.

Burghout, C. (2013, December). Workshop Rubrics op de studiemiddag Opleiding Pedagogiek, Fontys Tilburg.

Joosten-ten Brinke, D. (2013, 5 november). Betrouwbaarheid door formatief toetsen. Lezing op de Fontys toetsconferentie. Tilburg: Fontys. <http://fontys.mediamission.nl/Mediasite/Play/d48288d149294915be447ea523f340531d>.

Burghout, C. & Artis, J. (2013). Rubrics en schriftelijke feedback. Presentatie bij de inspiratiedag onderzoek Educatieve lectoraten.

Joosten-ten Brinke, D., Sluijsmans, D., Van der Vleuten, C., Arts, J., Jaspers, J., Burghout, C., Meusen-Beekman, K., Boshuizen, E. (2013, 29 mei). Waardevol onderwijs: Het effect van formatief toetsen op de kwaliteit van leren. Symposium bij de Onderwijs Research Dagen. Brussel: ORD.

Sluijsmans, D. M. A., Joosten-ten Brinke, D., & Van der Vleuten, C. (2013, 29 mei). Toetsen met leerwaarde. Presentatie bij de Onderwijs Research Dagen. Brussel: ORD.

Burghout, C. & Joosten-ten Brinke, D. (2013, 29 mei) De toegevoegde waarde van rubrics in formatief en summatief assessment. Presentatie bij de Onderwijs Research Dagen. Brussel: ORD.

Arts, J. G., Jaspers, M., & Joosten-ten Brinke, D. (2012, 29 mei). Analyse van schriftelijke feedback. Presentatie bij de Onderwijs Research Dagen. Brussel: ORD.

Joosten-ten Brinke, D. (2013, 26 maart). Het formatief gebruik van summatieve toetsen. Lezing bij de miniconferentie Formatief toetsen. Heerlen: Open universiteit.

Sluijsmans, D. & Joosten-ten Brinke, D. (2013, 26 maart). Wat vraagt formatief toetsen van de organisatie? Workshop bij de miniconferentie Formatief toetsen. Heerlen: Open universiteit.

Joosten-ten Brinke, D. (2012). Leerlingbetrokkenheid bij natuurkundeonderwijs. Lezing bij het Woudschotenconferentie, Noordwijkerhout: WND.

Joosten-ten Brinke, D., Sluijsmans, D., & Van der Vleuten, C. (2012). Which assessment features shape students' learning? A review study. Presentation at the Eapril conference, Jyväskylä, Finland. <http://hdl.handle.net/1820/4754>

- Burghout, C. (2012). Teachers' And Students' Perceptions Of The Additional Value Of The Use Of Rubrics In Both Formative And Summative Assessment. Presentatie bij de EAPRIL Jyväskylä, Finland.
- Burghout, C. & Joosten-ten Brinke, D. (2012, November). Inspireersessie Rubrics. Toetsconferentie: Fontys Lerarenopleiding Tilburg
- Joosten-ten Brinke, D. (2012). Alle docenten toetsbekwaam?!. Lezing bij de Fontys toetsconferentie. 6 november 2012. Tilburg: Fontys Hogescholen.
- Burghout, C. (2012, maart). We do Need Assessment for Learning. Keynote TED Istanbul Conference, Istanbul.
- Burghout, C. (2012, November). Assessment for Learning en Feedback. Lezing bij 'Assessments Minoren', NHL Opleiding CMD: Leeuwarden
- Joosten-ten Brinke, D. & Sluijsmans, D. (2012). Toekomstbestendig toetsen: effectief gebruik van formatief toetsen. 7 juni 2012. Workshop bij de EHON conferentie. Heerlen: Open universiteit/ Zuyd Hogeschool. <http://hdl.handle.net/1820/4753>
- Joosten-ten Brinke, D., Sluijsmans, D., & Van Eldik, S. (2012). Considered and competent assessment: A research on what teacher educators should know about assessment. 4 april 2012. Presentatie bij de ATEE conferentie. Coimbra: ATEE.
- Sluijsmans, D., Van Eldik, S. & Joosten-ten Brinke, D. (2012). Bewust en bekwaam toetsen en beoordelen: Wat vinden lerarenopleiders belangrijk? 7 februari 2012. Presentatie bij de conferentie voor lerarenopleiders. Antwerpen: Velov.
- Van Leuven, A., Lansu, W., & Joosten-ten Brinke, D. (2012). De kennistoets getoetst. 6 februari 2012. Presentatie bij de conferentie voor lerarenopleiders. Antwerpen: Velov.
- Joosten-ten Brinke, D., Sluijsmans, D. M. A. & Florack, A. (2011, 25 Novembre). How to improve teacher trainers' assessment knowledge and skills? Workshop presented at the Eapril conference, Nijmegen, the Netherlands. <http://hdl.handle.net/1820/3903>.
- Arts, J., Jaspers, M., & Joosten-ten Brinke, D. (2011, 24 Novembre). Written comments as a form of feedback; from prompting and imitation to deep learning. Workshop presented at the Eapril conference, Nijmegen, the Netherlands.
- Burghout, C., & Joosten-ten Brinke, D. (2011, 24 November). The Impact of Rubrics on Assessment Attitudes in Literature Studies. Workshop presented at the Eapril conference, Nijmegen, the Netherlands.
- Joosten-ten Brinke, D. (2011, 23 november). "Organisatie op scherp". Wat betekent de invoering van digitaal toetsen voor uw organisatie? Masterclass verzorgd tijdens de I&I conferentie, Lunteren: I&I conferentie. <http://hdl.handle.net/1820/3922>.
- Burghout, C. (2011, 16 September). The impact of Rubrics on Assessment Attitudes in Literature Studies. Lezing aan de Universiteit van Innsbruck.
- Verhoosel, J. (2011). Random Assessed Peer Assessment (RAPA). Presentatie bij de ElWier bijeenkomst. Hogeschool Arnhem en Nijmegen.
- Klatte, E. & Joosten-ten Brinke, D. (2011, 1 april). Didactiek en beoordelen bij CGO. Twee kanten van dezelfde medaille. Workshop gepresenteerd tijdens de conferentie 'Beroepsonderwijs op 1', Zeist. <http://hdl.handle.net/1820/3813>.

Sluijsmans, D. M. A. & Joosten-ten Brinke, D. (2011, 14 maart).

Bewust en bekwaam toetsen en beoordelen. Zicht op de toetskennis- en kunde van lerarenopleiders. Workshop gepresenteerd tijdens de Velon conferentie, Noordwijk.

<http://hdl.handle.net/1820/3816>

Joosten-ten Brinke, D. (2011, 1 februari). (on)mogelijkheden van digitale toetsen. Lunchlezing, FLOT, Tilburg. <http://hdl.handle.net/1820/3812>

Sluijsmans, D. M. A. & Joosten-ten Brinke, D. (2010, Novembre).

How can we define assessment quality: Exploring the quality pyramid of assessment. Workshop presented at the Eapril conference, Lisbon, Portugal.



Desirée Joosten-ten Brinke

Deze praktijkpublicatie laat de opbrengst zien van vier jaar praktijkonderzoek in het lectoraat 'Eigentijds toetsen en beoordelen' van Fontys Lerarenopleiding in Tilburg.

De kenniskringleden presenteren onderzoek naar hun eigen toetspraktijk en een masteronderzoek. Tevens is een hoofdstuk gewijd aan het onderzoek 'Bewust en bekwaam toetsen' dat de kenniskring samen met drie andere instituten in een SURF-project heeft uitgevoerd.

Fontys Lerarenopleiding Tilburg

Fontys Hogescholen

Prof. Goossenslaan 1-01 (Mollergebouw)

Gebouw P1

5022 DM Tilburg

Tel.: 0877 883 375

www.fontys.nl/lectoraten/eigentijdsbeoordelen